Patents and Patent Races. Do We Need Them? How Should We Behave?

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

doctor rerum politicarum

(Doktor der Wirtschaftswissenschaft)

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

M.Sc., Cristian Stefan

Präsidentin der Humboldt-Universität zu Berlin:

Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Wirtschaftswissenschaftlichen Fakultät:

Prof. Dr. Christian D. Schade

Gutachter:
1. Prof. Dr. Christian D. Schade
2. Prof. Dr. Ulrich Kamecke

Tag des Kolloquiums: 02.11.2016

PATENTS AND PATENT RACES.

DO WE NEED THEM?

HOW SHOULD WE BEHAVE?

Table of Contents

CHAPTER 6

Index of Tables

Index of Figures

Table of Abbreviations

Abs.	Absatz (paragraph)
a.k.a.	also known as
A/N	author’s note
e.g.	exempli gratia
EPO	European Patent Office
FDA	Food and Drug Administration
GATT	General Agreement on Tariffs and Trade
IMD	incrementally modified drug
IP	intellectual property
IPC	International Patent Classification
NIHCM	National Institute for Health Care Management
NME	new molecular entity
OECD	Organization for Economic Cooperation and Development
PatG	Patentgesetz (German patent law)
PC	Personal Computer
PCT	Patent Cooperation Treaty
PR	Public Relations
R&D	research and development
UK	United Kingdom
US, USA	United States of America
USPTO	United States Patent and Trademark Office

CHAPTER 1

Introduction:
How Patents Matter to Business, Economics and Society

"If patent law had been applied to novels in the 1880s, great books would not have been written."

The Guardian (23.06.2005)

"Some of the most innovative industries of the last forty years – software, computers, and semi-conductors – have historically had weak patent protection and have experienced rapid imitation of their products."

Bessen and Maskin (2009)

"On the patent front, more time and energy seems to be spent on nuisance and defensive patenting of the obvious or well-known than is spent on actually innovating new ideas."

Boldrin and Levine (2002)

1. Introduction

The year is 2012. On an August morning, in front of Apple Computer's headquarters in California arrive 30 large trucks, each fully loaded with five-cent coins, and totaling a billion dollars. Apple's chief executive Tim Cook is called by Samsung and told this is the way Samsung wants to pay the fine to Apple after losing the latest patent infringement trial on smartphones and tablet computers. The story of the trucks, which was very popular on the Internet, with news articles, pictures and videos dedicated to making it more credible, is of course false (although the billion-dollar fine was real). One would need more than 2,700 trucks to carry a billion dollars in nickels; the coins would weigh 100,000 tons and would necessitate the entire amount of nickels in circulation in the US (The Guardian, 29.08.2012). Even though the story was a hoax, it marked a peak of the so-called "smartphone patent wars", which continually made news headlines during those years.

The smartphone patent wars, which have been ongoing for years, were largely fought on a growing market and involved countless patent litigation trials between giants of the Internet, computer and mobile phone industries, with Sony, Google, Apple, Samsung, Microsoft, Nokia, Motorola and HTC as the most prominent examples. These wars were so complex that even Eric Schmidt, Google's chairman, said he didn't "understand all the details" of the smartphone patent fights (The Guardian, 22.10.2012). One of the reasons for the confusion is that smartphones "combine technologies from different industries, most of them patented. Given such complexity, sorting out who owns what requires time and a phalanx of lawyers" (The Economist, 21.10.2010). I discuss this peculiar phenomenon in my dissertation; it has even won itself a name: "patent thickets". To make matters worse, the smartphone wars became all the more complicated when "patent trolls" got involved. These are firms "that buy patents not in order to make products, but to sue others for allegedly infringing them" (idem). "Patent trolls" are generally seen as a perverse result of the patent system, and it is unfortunate that they have also gained a name. Patents, which were originally intended to spur innovation, seem meanwhile to have drifted from their purpose in many ways. A large portion of this dissertation is dedicated to the analysis of such distortions.

Among the smartphone wars, the battle between Apple and Samsung particularly stood out because, while fighting over patents, "the two companies are close business partners, with Apple as one of the biggest buyers of chips, screens and other components that Samsung manufactures" (The Wall Street Journal, 22.04.2011). This makes the situation somewhat absurd; after all, the two firms could have agreed on a business solution and would have spared themselves the enormous costs of going to court. Indeed, the literature has signaled many concerns that patent battles redirect significant resources away from innovative activity into wasteful legal suits.

Even more absurd, the one-billion-dollar battle from August 2012 was won on "design patents", not on regular patents. That is, it was won on form, not function. Apple had registered designs for elements including the *rounded corners* of its iPhone and iPad, and argued these corners were "visually distinctive features" of their products that Samsung infringed upon (The Guardian, 22.10.2012). As we will see throughout the dissertation, such patents which lack real novelty (often under the name of "trivial patents") may severely jeopardize the chances of subsequent innovation and therefore harm society and the economy. In the Apple-Samsung patent war, Samsung was even banned from selling its tablet in the US, at the detriment of consumers who wanted to buy the product. After receiving the billion-dollar fine, Samsung declared that the "verdict should not be viewed as a win for Apple, but as a loss for the American consumer. It will lead to fewer choices, less innovation, and potentially higher prices. It is unfortunate that patent law can be manipulated to give one company a monopoly over rectangles with rounded corners, or technology that is being improved every day by Samsung and other companies" (The Guardian, 25.08.2012).

Ironically, the entire battle resulted in some unintended consequences for Apple. Initially, "many people had no idea that Samsung had a tablet offering until they began hearing news reports about Apple seeking to ban it" (idem), and people started buying Samsung afterwards. Eventually, Apple hurt its own reputation through the patent war, and turned Samsung into the second largest seller of smartphones worldwide.

The example of the smartphone wars was meant to point out some patent-related issues that we will encounter again throughout this dissertation. I believe that patents have generally recognized virtues and essential features for firms, investors, the economy and society at large. Among others, patents are meant to protect the efforts of innovators against easy imitation by competitors; therefore, patents' role is indispensable. On the other hand, patents can bring a myriad of unwanted and unforeseeable consequences to the involved institutions. Thus, it is a very difficult endeavor for researchers, policy makers and regulators to design an "optimal" level of patent protection. Finding a balance between too strong and too weak patent protection will often depend on many factors: the industry structure, product characteristics, the type of innovation as well as other forces. In my dissertation, I bring a theoretical and experimental research contribution by studying the *status quo* on patents and patent races, shedding light on critical issues and discussing potential improvements of the patent incentive structure.

1.1. Business and Economics Aspects of Patents

1.1.1. A Brief Patent Definition

In this subchapter, I give a definition of patent protection according to German patent law (PatG) and I mention further features of patents. Protection is achieved by the fact that, after the granting of a patent, third parties are no longer allowed to offer, sell, use or produce an invention, except if the patent holder allows them that through licenses (§§ 9, 15, 23 PatG). This exclusive protection right can only be awarded for a maximum of 20 years, but still not any invention can be patented. The following conditions must be fulfilled in order for exclusive protection to be granted: the invention must be new, based on an inventive activity and commercially usable (§ 1 Abs. 1 PatG). To be considered new, the invention may not belong to the state of the art (§ 3 Abs. 1 PatG). To the state of the art belongs all knowledge, written or spoken, which was available to the public before the day of the patent application. The product or process is considered an inventive activity if the description of the inventive step is not obvious to an expert professional (§ 4 Abs. 1 PatG). The last condition for patenting is that the invention must be produced or be usable in an industrial field of application (§ 5 Abs. 1 PatG). There are also many products and ideas that are not patentable, for instance discoveries, mathematic theories and scientific methods; aesthetic conceptions; plans, rules and procedures for intellectual activities; computer programs which do not contain technical instructions and the reproduction of information (§ 1 Abs. 3 PatG).

A patent document typically contains: “bibliographic data (providing information on the applicant, inventor, technological classes, references to the existing prior art, etc.), a specification or description of the invention, a set of claims (what the patentee is claiming exclusive rights on, a.k.a. the patent’s scope of protection), and finally – but optionally – some illustrations supporting the specification and claims in the form of drawings, listings, gene sequences, etc.” (van Zeebroeck et al., 2009). First and foremost, the purpose of a patent is to protect a product or a method against imitators. But patents are used for a whole range of different tasks. They may be used to block competitors, to win technological reputation, to produce internal incentives or as a means of exchange (cross-licensing, co-operations, access to the capital market, revenues from licensing) (Blind et al., 2003). Many of these aspects are discussed in the following.

1.1.2. Arguments For and Against Patents

The traditional economic argument for intellectual property protection appears in the pioneering works by Arrow (1962) and Nordhaus (1969): innovative activity leads to the production of knowledge, but, unlike other goods, knowledge has an inherently non-rival character. This means that once an invention is known, everyone can use it without incurring additional R&D costs, which induces market failure and destroys initial incentives to innovate. Patents are supposed to reverse that effect and provide incentives for innovation. Fershtman and Markovich (2010) write that “the conventional wisdom is that by protecting innovators from imitation we encourage R&D investment and promote innovation”. However, these authors continue by saying that “recently this rationale has been challenged” (they are referring to empirical evidence from the software and computer industries, which experienced the most rapid innovation during periods of weak patent protection; I discuss some of these findings in Chapter 1.1.4).

Among many researchers, Boldrin and Levine (2002) challenge the conventional view on patents. They show that intellectual property has two components: the right to own and sell ideas, which they view as essential, and the right to control the use of those ideas, which they regard as economically dangerous, because it results in an “intellectual monopoly” with very high social cost. Boldrin and Levine (2002) do admit that “if strong property rights provide good incentives for the production of potatoes, they must also provide good incentives for the production of ideas", but they point out to the very different ways the ownership of products versus ideas is treated in practice: "When you buy a potato you can eat it, throw it away, plant it or make it into a sculpture. Current law allows producers of CDs and books to take this freedom away from you. When you buy a potato you can use the "idea" of a potato embodied in it to make better potatoes or to invent French fries. Current law allows producers of computer software or medical drugs to take this freedom away from you. It is against this distorted extension of intellectual property rights that we argue".

In their article, Encaoua et al. (2006) summarize the most important virtues and drawbacks of the patent system. In this paragraph, I discuss the four virtues provided by Encaoua et al.; I closely follow the reasoning of these authors but I also add my own interpretation of their statements:

- the first virtue is that the government gives some temporary exclusive property rights to inventors but in exchange delegates the R&D investment decision and therefore the responsibility as well as risk of the R&D cost fall into the hands of private inventors. This is a wise decision, because individual agents have much better information on the benefits and costs of R&D than the government does. Another positive consequence of delegating is that the potential problem of moral hazard on the side of researchers is alleviated (Encaoua et al., 2006). If, instead of delegating, the government chose to subsidize R&D investment, inventors could use the subsidy for private purposes and pretend to be unsuccessful in their R&D endeavour (which would be a plausible statement, due to the highly uncertain nature of R&D);
- the second virtue of patents is that the assignment of costs falls on the users of the products, and not on taxpayers (idem). That is, the R&D cost (per unit) is included in the price of the product and affects buyers of the product specifically and not the general public who might have no interest or connection to a certain patented product;
- the third is that the government may instill a patent system without requiring "sensible economic information that is only privately known, such as R&D cost and private value of the invention, thus avoiding adverse selection problems" (idem). In this context,

adverse selection would mean that, without patents and in the absence of information, the government might choose to finance a firm that is either cost or revenue inefficient, or both. Instead, through the patent system, the effective selection of R&D projects is done by the market: innovative firms weigh the private values of their inventions against the costs of patenting and decide whether to invest or not;
- the fourth virtue is that patents require the disclosure of information, which helps the further diffusion of knowledge (idem) and therefore subsequent innovation.

Supporting these virtues, a prestigious publication by Kultti et al. (2007) also suggests that the patent system is able to simultaneously stimulate information disclosure, innovation and welfare.

On the other hand, Encaoua et al. (2006) also find five drawbacks of patents and note that there might be more. Indeed, throughout this dissertation I found many other weaknesses of patents and discussed them in detail.

- The first weakness identified by Encaoua et al. is that patents induce "the classical deadweight loss that results from inefficient monopoly pricing: not all consumers valuing goods above their marginal cost can buy them".
- Second, the prize obtained from a patented good (profits, market share) is not directly linked to the R&D effort necessary to develop it (idem). In fact, Chapter 3.1.2 shows that in recent years firms almost completely decoupled their patenting and their R&D activities; besides, Chapter 2.2.2 finds no relationship between patent protection and innovative activity.
- The third drawback signaled by the above authors is that firms racing for a patent may be subject to a wasteful duplication of R&D costs. This thought worried many patent researchers and is very present in the patent literature. However, an empirical study by Cockburn and Henderson (1994) shows that such a wasteful behavior does not occur even in the most patent-intensive industries; I discuss the topic in Chapter 4.4.
- The fourth drawback is derived from theoretical work by Eswaran and Gallini (1996): patents are much more likely to create substitute products instead of complementary goods and this results in coordination problems on the market. To my mind, this happens

because, if more substitutes are being produced, competition between existing industry players intensifies and the market does not settle to accommodate firms who harmonize their activities by producing complementary products.

- The fifth weakness is related to opportunity costs: patenting requires a significant "amount of financial resources that are diverted from the innovation process itself" (Encaoua et al., 2006).

There are countless arguments for and against patents, and I discuss most of them throughout the dissertation. Still, after scanning the literature, I can assert that most papers do have one element in common: they all agree that too weak intellectual protection rights might lead to an under-provision of R&D, while too strong protection rights might lead to an excessive monopoly distortion in the form of a deadweight loss that has to be incurred by society, and potentially to a slow-down in the speed of innovation. Therefore, any measure to strengthen or weaken patent protection involves a series of difficult tradeoffs. In the words of Encaoua et al. (2006): "patents appear to be a second best solution. The first best, characterized by a socially desirable level of innovation without market power and with global diffusion, appears to be unreachable".

1.1.3. The Importance of Patents For Firms

According to a survey from 1999 performed by the Fraunhofer Technology Development Group (Fraunhofer TEG 1999), German companies use about 38% of their patents for products that they sell, while 43% of patents are not used at all. Although at first glance the proportion of unused patents seems rather large, it may be a natural consequence of the fact that firms make *ex ante* decisions to patent. That is, firms decide to patent before they know the real market benefits of the patented product. From a Bayesian information point of view, a firm decides to patent with incomplete information, based on prior beliefs about the future market rewards from patenting. After patenting, the firm receives market feedback and updates its prior beliefs *ex post*. In a sense, patenting activities are a necessary cost, similar to advertising activities; I reiterate this point at the end of Chapter 1.2.1. Fraunhofer TEG (1999) also reports that the remaining 19% of the held patents are so-called "strategic patents", which are either sold to other companies or used to delay the market entry of other products, because competitors are forced to research around the strategic patent in order to avoid infringement. Patents serve further purposes, different from the exclusion of competitors from the market: firms use patents to obtain negotiation power "for cross-licensing agreements, as a signalling mechanism for shareholders, banks, venture capitalists, competitors or customers" (Encaoua et al., 2006). In this subchapter I limit my analysis to the importance of patents as protection mechanism for firms' innovations.

Harhoff (1997) reports on the analysis of a dataset entitled Mannheim Innovation Panel, which was collected by the German government with the purpose of providing a comprehensive description of German firms' innovative activities. Harhoff (1997) used a cross-sectional portion of the dataset, from the year 1993, and researched a number of 1,413 West-German companies on the estimated effectiveness of different protection mechanisms, through which they appropriate the returns from their innovative activities. Tables 1.1 and 1.2 provide results of the investigation for product and process innovations. For both types of innovations firms rely largely on alternative protection mechanisms such as personnel binding, first mover advantages, complexity and secrecy, while intellectual property comes last with patents, copyright and registered trademarks. 42 % of companies consider patents effective as protection mechanism for product innovations, while for process innovations the figure is lower: 25.2 %. Similar general estimates for the importance of patents as protection mechanism were obtained by Levin et al. (1987), who performed a survey of high-level R&D executives in the USA. They also found that "lead-time advantage" and "moving down the learning curve quickly" were rated as a more effective protection mechanism than patents. Finally, patents have an important drawback for firms, because they require the publication of the knowledge embedded in the patent. Horstmann et al. (1985) emphasize that this may lead to a danger of knowledge drain for the patenting firms.

1.1.4. Sequential Innovation

The conventional view on patent protection (as described in Chapter 1.1.2) is turned upside down in the ground-breaking paper by Bessen and Maskin (2009). They point out to empirical evidence showing that some of the most innovative industries towards the end of the 20th century – software, computers and semiconductors – had historically been under weak patent

Table 1.1. The importance of protection mechanisms for product innovations

Protection mechanism	**Percent of firms which consider this mechanism important or very important**
Long-term binding of qualified personnel	83.7 %
Lead-time advantage	82.6 %
Product complexity	55.3 %
Secrecy	53.4 %
Patents	42.0 %
Copyright and registered trademarks	25.3 %
No. of firms	1,413

Source: Harhoff (1997), based on cross-sectional data from the 1993 Mannheim Innovation Panel.

Table 1.2. The importance of protection mechanisms for process innovations

Protection mechanism	**Percent of firms which consider this mechanism important or very important**
Long-term binding of qualified personnel	79.6 %
Lead-time advantage	79.6 %
Process complexity	52.2 %
Secrecy	49.8 %
Patents	25.2 %
Copyright and registered trademarks	14.8 %
No. of firms	1,413

Source: Harhoff (1997), based on cross-sectional data from the 1993 Mannheim Innovation Panel.

protection. In the US, patent enforcement for semiconductors and computers first came into place through the Federal Circuit Court in 1982 (idem). Before this measure, the industry was patent free. Software had been excluded from patent protection for decades, until a series of court decisions in the mid-1980s and 1990s significantly increased the strength of software patents, which Bessen and Maskin (2009) consider “a revealing natural experiment”. This experiment showed that, instead of generating a burst of innovative activity, the firms which acquired the software patents actually reduced their R&D spending as percentage of sales (Bessen and Hunt, 2004). Bessen and Maskin (2009) state that “there is nothing paradoxical about this outcome”. In their view, for industries such as software or computers imitation may *promote* innovation, while strong patents might in fact *inhibit* it, contrary to the conventional view. Bessen and Maskin argue that those industries which seem to be doing better without patent protection share one common element: the presence of sequential and complementary innovations. They define an innovation as sequential if “each successive innovation builds on the preceding one”. They illustrate this with the example of spreadsheet software, where Microsoft’s Excel built on Lotus, and Lotus built on VisiCalc. Furthermore, Bessen and Maskin define an innovation as complementary if “each potential innovator takes a different research line and thereby enhances the overall probability that a particular goal is reached within a given time”. The authors ascertain that the complementarity of the different research lines taken to voice-recognition software increased the speed of this innovation.

In a world of innovations which are both sequential and complementary, the conventional reasoning about patents’ role in the promotion of innovation seems to be reversed. Bessen and Maskin give reasons for this counter-intuitive result: imitation of a discovery is socially desirable in such a world, because it helps the imitator develop further inventions, and since the imitator may have valuable ideas not accessible to the original inventor, the overall speed of innovation may be enhanced. What is more, with sequential innovations even the inventor herself could find it advantageous if others imitate and compete against her. Even though imitation reduces the inventor’s current profit, it increases the probability of follow-on innovations, which in turn improves the inventor’s future profits. Bessen and Maskin even bring relevant examples of firms which appeared to encourage imitation. In 1981, IBM announced its first personal computer;

Apple Computer, which was then the industry leader, responded with a full-page newspaper ad entitled “Welcome, IBM. Seriously”. Cisco, among others, regularly contributes patented technology to the industry standard and allows any market entrant to produce competing products. IBM and other firms donated a number of patents to be used for free by developers of open source. Bessen and Maskin do admit that the conclusions from their paper would fail to work if the cost of market entry were equal to zero, or if imitation were infinitely fast. However, in real-world settings market entry requires significant investments in specialized capital, and entry does not occur instantly, so that the original innovator typically enjoys a temporary first-mover advantage. Thus, Bessen and Maskin’s model cannot be overthrown by rejecting its main assumptions.

Chapter 1.1 reviewed the conventional view on patent protection and showed some of the ways in which this view has been challenged by more recent research. While patents do have their virtues both on the macroeconomic and on firm level, they also bring a number of drawbacks to the economy as well as society. Moreover, there are very important settings, such as sequential and complementary innovations, in which patents seem to do more harm than good; other such situations can be found throughout the dissertation. In the next chapter I concentrate on two major disadvantages brought by the patent system in practice: the explosion in patenting at the detriment of patent quality and the emergence of a dangerous phenomenon under the name of patent thickets.

1.2. Distortions Induced By Patents

1.2.1. Quantity Before Quality

Van Zeebroeck et al. (2009, p. 1011) write that the intellectual property (IP) strategy of firms in some industries recently moved from being static to more active. That is, companies no longer use patents just for the leveraging of exclusion rights. Instead, they started using them for licensing and other strategic purposes. More importantly, firms changed their patent strategy from a “single patent” view to a “portfolio management” view. Some large firms have a strong tendency to patent their inventions. “IBM for instance has used the fact that it is the largest patentee in the USA as a marketing tool for several years and Microsoft has recruited an IP officer (from IBM) to develop its patenting strategy with a view to outperforming IBM in this respect" (idem). Unfortunately, the shift from the single patent strategy to the patent portfolio strategy resulted in a deterioration of the quality of patents, an explosion of patent applications and a world in which the size and strength of the patent portfolio are more important than the original purpose of patents, which is to reflect real inventive activity. Many researchers have expressed concern about this "more is better" strategy, which perverts patenting incentives, places patent offices under pressure and compromises the reliability of the patent system by creating patent thickets that no one can look through anymore.

Even more paradoxically, the boom of patent applications was not coupled with an increase in innovative activities. For instance, an industrial inquiry among electronic companies done by Taylor and Silberston (1973) showed that firms uniformly agreed that patents have very little or no impact on R&D investment decisions and on the industry competition. Patent expert and lawyer Jan Tönnies states that to interpret the increase of patent applications as innovative progress makes as little sense as interpreting an increase in the use of prisons as a success in combating criminality (Wissensallmende, 2009). A detailed discussion on the effects of the unprecedented surge in patent applications is provided in Chapter 3.

Adding to the discussion on the diminishing quality of patents, Taylor and Silberston (1973) point out that a large proportion of patents in electronics and electrical engineering is of "very suspect validity". They report that up to 90% of these patents are questionable due to a lack of clear consensus about their novelty. Different than with chemical research, where it is clear whether a chemical compound is new or not, with electronics many innovations represent very small technical improvements or rearrangements of existing components and circuits, which are not necessarily patentable (idem). Taylor and Silberston (1973) show that even basic patent rights held by the American company Western Electric for the transistor could easily be attacked by UK firms who refused to pay royalties for the use of transistors. Similar findings to Taylor and Silberston are brought by other researchers. Bessen und Hunt (2004) are very critical about the results of a strengthening of patentability which took place in the US software industry. According to their econometric study on the patent-R&D ratio (which I also discuss in

Chapter 3.1.2), Bessen and Hunt (2004) prove that stronger patent protection for software did not lead to more innovative activity and R&D. Instead, it allowed firms to appropriate portions of the revenues obtained by successful inventions through the use of "trivial patents" – patents lacking real novelty and granted after very limited R&D efforts. Lanjouw and Schankerman (2004) corroborate Bessen and Hunt's findings by looking at the patent-R&D relationship within a traditional invention production function and taking into account patent quality (they build a quality index for patents). Their results confirm that patents of low quality were responsible for the increase in the patent-R&D quotient. Overall, the increase in patent applications, coupled with a questionable patent quality, generates an uncertainty which compromises the reliability of the patent system altogether (van Zeebroeck et al., 2008). I discuss this in Chapter 3.3.

Yet, in spite of patents' allegedly falling quality, since the 1970s the number of patents held by a firm has been one of the criteria for rankings at the stock exchange: a firm with more patents was considered more innovative and more valuable. This incentivized firms to patent more, even though they could not bring to market many of their patented developments (Wissensallmende, 2009, based on data from the Fraunhofer TEG, 1999). According to popular science, firms ended up in an arms' race with patents: if one firm patents, all other firms have to patent as well! (idem). In game-theoretical terms, firms can be seen as caught in a Prisoner's Dilemma where patenting is the non-cooperative strategy. However, the same point could be made for advertising. It is generally recognized that advertising, as well as the costs it brings, is an essential activity performed by firms. I regard patenting activities from a similar perspective: they are costly and their success is uncertain, yet all firms perform such activities, because they are necessary for survival.

1.2.2. Patent Thickets

The rise of patenting in recent decades was coupled with an increase in technological complexity and these combined factors gave birth to a dangerous phenomenon labeled as "patent thickets". Bessen (2003) writes that the traditional patent race literature assumes a "one-to-one correspondence" between a product and a patent. In reality, technologies are complex and property rights for a product are shared by several firms through a number of patents. More often than not, a product may be based on hundreds of patents, and this generates patent thickets, "a dense web of overlapping intellectual property rights that a company must hack its way through in order to actually commercialize new technology" (Shapiro, 2001). Bessen (2003) as well as Siebert and von Graevenitz (2010) explain that patent thickets are a result of firms' patent portfolio strategies: the more patents a firm has related to a certain product, the greater its probability of winning in patent infringement suits, and thus the stronger the firm's bargaining power and the larger its "freedom to operate". But patent thickets have a dark side: if an innovator must contact and negotiate with a large number of patent holders, her transaction costs may become prohibitively high (Heller and Eisenberg, 1998). Shapiro (2001) quotes the famous statement by Sir Isaac Newton that every scientist "stands on the shoulders of giants" and puts it in relation to patent thickets: "today, most basic and applied researchers are effectively standing on top of a huge pyramid, not just one set of shoulders". Shapiro points out that the pyramid may reach "far greater heights" than any individual alone. However, this will not happen if a researcher who wants to place a new block on the top of the pyramid must, in Shapiro's words, "gain the permission of each person who previously placed a block in the pyramid". In patent terminology, such persons are the owners of blocking patents, and they might be able to prohibit innovators from using the invention. Consequently, patent thickets may drastically slow down the commercialization of new discoveries and ultimately decrease the speed of innovation, which is exactly the opposite of what patents are intended to do (idem).

Shapiro (2001) emphasizes that the patent system is creating patent thickets and severe transaction costs in several key industries: semiconductors, biotechnology, software and the Internet. In Chapter 1.1.4, I showed that such industries (where sequential innovation is of essence) are particularly prone to incur strong negative effects of patenting. Shapiro (2001) supports this view: "With cumulative innovation and multiple blocking patents, stronger patent rights can have the perverse effect of stifling, not encouraging, innovation". Thus, patent thickets are attacking our newest, most sensitive and essential industries. The literature (particularly Shapiro, 2001, Bessen, 2003 and Siebert and von Graevenitz, 2010) does come up with some solutions to the thicket problem: cross licenses (where two companies grant each other the right to use the other's patents) and patent pools (where several patent holders agree to jointly license an entire group of patents and share the revenues). However, Bessen (2003) expresses doubts about the effectiveness of such institutions and points out that the deepest problem causing patent thickets is rooted in low patent standards. According to Bessen, when patentability standards are high, "portfolio building is costly" and firms are more likely to embark in activities that rely less on patents. Conversely, low patent standards result in aggressive portfolio building, socially wasteful behavior and even reduced R&D incentives (idem). Overall, many of the distortions discussed in this chapter could be alleviated by reversing the tendency towards lower patent standards which could be observed during the last decades and returning to a more reliable regime with patents of high quality, based on real novelty and significant inventive effort.

1.3. Patents, Society and Ethics

Apart from patents' economic use (where we could see that patents have their virtues but also create many distortions), there are important societal and ethical considerations related to intellectual property, especially in some sensitive industries where people's health and lives are at stake. In her excellent book, Naomi Klein (2009) raises the alarm about profit-seeking firms attempting to make fortunes out of putting "a patent and a price tag on life-forms and natural resources never dreamed of as commodities – seeds, genes, carbon in the earth's atmosphere". Patents are even more ethically questionable in public health issues. Klein (2009) refers to the outbreak of polio (a crippling and often fatal children's disease) in the US during the 1950s, when "the ethics of disease profiteering were hotly debated". With around 60,000 cases of polio, the "search for a cure was frantic" (idem). Finally, a scientist at the University of Pittsburgh, Jonas Salk, found the cure, developed the first polio vaccine in 1952 and decided not to patent this lifesaving treatment (idem). When asked about his decision, Salk answered "There is no patent. Could you patent the sun?" (Oshinsky, 2005). Such examples of highly ethical behavior should be more common in our society. Unfortunately, in practice one finds some bad apples; for instance, Klein (2009) points out that Gilead Sciences, a company which owns patents on four AIDS treatments, puts a lot of effort in "trying to block the distribution of cheaper generic versions of its lifesaving drugs in the developing world", even though some of Gilead's "key medicines were developed on grants funded by taxpayers”. From a societal point of view, Gilead makes a double mistake. This example raises serious ethical concerns about patent protection for medical drugs, where the patent may prevent saving lives. Due to the importance of this debate, I provide a detailed analysis of patents for pharmaceuticals in Chapter 4. However, there are many other fields where society at large is against the use of patents: genetics, nanotechnology, stem cell and biomedical research are some of the most prominent areas of concern I have identified in the literature. In 2013, the US Supreme Court has already decided in the case of Myriad Genetics, a company holding patents on two genes linked to breast cancer, that human genes are not patentable (The Economist, 13.01.2013), a decision that will hopefully also stay valid in the future.

Furthermore, the general public has started to be more and more informed and critical about trivial patents, which do not fulfill the usual standards for novelty and non-obviousness. Shapiro (2001) writes that “even while a consensus has emerged that innovation is the main driver of economic growth, we are witnessing somewhat of a backlash against the patent system as it is currently operating”. Shapiro refers to especially unpopular patents such as Amazon's patent on a one click online shopping system, which was asserted against Barnes & Noble and caused a public scandal. The Guardian (22.10.2012) criticizes Apple’s “slide to unlock” system for the iPhone as another trivial patent where potential competitors had trouble inventing around the patent. For instance, “a touch interface with a thin line along which you slide a ball to unlock the screen” could have been judged as infringing Apple’s patent (idem). Eventually, Google’s Android system found a way to circumvent the “slide to unlock” patent by letting users pull a lock icon in any direction without using a line to hint in which direction the lock should be dragged (idem). While The Guardian (22.10.2012) sees most of the patent disputes as “boring” and “depressing”, I regard them as dangerous to society, because trivial patents such as the ones described above create obstacles for subsequent innovators and therefore may lead to a slow-down or even blockage of innovation.

There are several further societal concerns about the use of patents, out of which I only mention one more, related to fairness issues. The Telegraph (16.12.2010) writes that patents are “practically worthless” for small companies, because challenging a patent can only be done through very "lengthy and expensive" court processes, which small companies cannot afford. Thus, the patent system allows big firms with "deep pockets" to copy innovations of smaller firms (thereby infringing IP rights) and then to avoid any legal challenge by assembling expensive legal teams that smaller competitors cannot pay for (idem).

In the following chapter, I present the structure of the dissertation and I highlight its main research and discussion topics.

1.4. Structure of the Dissertation

In Chapter 2.1 I provide and attempt to synthesize a series of scientific definitions for the three instruments of patent regulation which I have identified in the literature: patent breadth, patent height and patent length. Through such instruments, patent regulators are able to influence the appropriability of market benefits for producers of innovations and therefore these tools are of crucial importance to the level of innovativeness on any market. However, Chapter 2.2 shows that any attempt by regulators to find an optimal level of patent protection faces many difficult tradeoffs and even the concept of "optimality" is not easy to define. I discuss "optimality" of patent and patent race regulation with respect to three aspects, corresponding to the three subchapters included in Chapter 2.2. In Chapter 2.2.1 I analyze the welfare tradeoffs which the regulation of various patent instruments leads to, especially the tradeoffs between patent length and breadth; I include some of the important models which were developed in the literature. In Chapter 2.2.2 I look at the effects of patent protection on innovativeness by discussing changes in patent policy in different countries and industries (considered "unusual natural experiments"). Chapter 2.2.3 presents three papers which deal with optimal rules for patent races.

Chapter 3 discusses the dramatic increase in the number and size of patent applications around the world, which raises the alarm about patent quality standards. The chapter is structured in three parts: Chapter 3.1 describes the *status quo*, Chapter 3.2 identifies reasons for the surge in patent voluminosity, while Chapter 3.3 presents the effects of and possible solutions to the patent boom. I start Chapter 3.1.1 with scientific definitions of patent claims and their relationship to the number of pages pro patent. In Chapter 3.1.2 I describe the dynamics of patent growth, which became particularly dramatic after the 1980s; I also relate the development of patent voluminosity to R&D activities and try to show a link between them; this link is important because it leads us to the question whether firms patent out of "patenting greed" or due to structural changes in the economy. Chapter 3.2 shows that in fact both the "greed" and the structural changes might play an important role for the surge in applications. This chapter identifies five factors for the patent growth, which are discussed in the five subchapters 3.2.1 to 3.2.5: administrative fees of patent offices; internationalization of applications through the Patent Cooperation Treaty; national practices, particularly the differences between the US and Continental Europe; the increase in technological complexity, which requires more words to be described; the emergence of young and very complex sectors with less standardized vocabulary. In Chapter 3.3 I present the effects of the patent boom on patent offices, on the quality standards of patents and on the reliability of the patent system altogether. Towards the end of this chapter I sketch potential solutions from the literature, especially through the fee structure.

Chapter 4 analyzes the role of patents in the pharmaceutical industry. Chapter 4.1 starts by pointing out to the pharma industry being the most profitable of all US industries during the half of the 20th Century and to a strengthening of patent protection which seemed to encourage the emergence of such profits. I also discuss the role of the enormous development costs which are a factor influencing patent protection. In Chapter 4.2 I present the hot debate from the literature on the topic of innovativeness in the pharma industry; here, I introduce arguments and figures for and against stronger patent protection. In Chapter 4.3 I show how the competition for "follow on" or "me too" drugs has dramatically increased towards the end of the 20th Century and I raise again a question from Chapter 4.2 about "true innovators", which might need stronger patent protection against "copycats" (or imitators). In Chapter 4.4 I shed light on the topic of patent races for pharmaceuticals: I present firms' racing behavior and empirical evidence for such races. Here, the main research question is whether firms engage in cut-throat competition and wasteful duplication of R&D costs, or whether they settle to share the market amicably. Chapter 4.5 discusses the role of generics and whether they fulfill their role of conquering the market and driving down prices after the patent of the original drug expires. In Chapter 4.6 I wrap up my conclusions and recommendations about patents in the pharmaceutical industry.

Chapter 5 introduces the theoretical and experimental literature on patent races. The chapter is divided into two parts: in Chapter 5.1 I analyze theoretical models of patent races, while Chapter 5.2 presents experimental evidence and partially relates it to the theoretical literature. Chapter 5.1.1 reveals the “traditional” patent race literature and points out to the many contradictions encountered between the papers and models developed here. This chapter includes the literature from the time span between the late 1960s and the early 1980s. During the ‘80s, a shift towards multistage two-player patent races can be observed, and this is the topic of the following Chapter 5.1.2. In Chapter 5.1.3 I summarize the main results of the excellent literature review on patent races, performed by Reinganum (1989). She was able to wrap up the literature and synthesize it into 16 assumptions and 40 propositions, which are of great theoretical value. In Chapter 5.1.4 I present further ramifications of the patent race literature, most notably perpetual (or endless) races and the optimal design of patent races. Chapter 5.1.5 wraps up the findings from the entire Chapter 5.1. The next chapter (5.2) is separated into two subchapters. In Chapter 5.2.1 I look at experimental studies which took models from the patent race literature as their theoretical basis, and I compare results from theory with results from experiments. Differently, in Chapter 5.2.2 I describe patent race experiments which did not root their theory in the standard patent race literature but instead developed their own theory (based on established theoretical constructs) and their own experiments.

In Chapter 6 I extensively discuss my experimental study on patent races. The chapter consists of four parts, represented by Chapters 6.1 to 6.4, in which I analyze the following aspects: building up the game-theoretic patent race model; building the experiment; econometric analysis and results; conclusions. Chapter 6.1.1 starts with a selection of Nalebuff’s (1988) sailing race which later is adapted to become a patent race model; here, I also highlight my contribution to the existing literature. Chapter 6.1.2 builds up a novel patent race model by adding parameters to Nalebuff’s puzzle, parameters which are essential for the modeling of an R&D race under a time framework. In this subchapter I construct the generalized decision tree based on game-theoretic considerations, I define the model parameters and I generate the payoff matrix of the game. In Chapter 6.1.3 I perform an algebraic equilibrium analysis, I discard equilibria which may never exist and I formally point out to Nash equilibria which may exist, given certain parameters. The next chapter, 6.1.4, employs a multi-dimensional graphical solution based on several parameter sets. In Chapter 6.1.5 I select the appropriate parameters for an experimental study based on realistic interval definitions and interactions between parameters. In Chapter 6.2.1 I briefly point out to the importance of experiments as research methods. Chapter 6.2.2 presents the experimental treatments and the way the experiment was implemented. In Chapter 6.2.3 I formulate two propositions derived from the framework of perfect rationality and two behavioral hypotheses based on the framework of bounded rationality. Chapter 6.3 employs statistical methods to show the results of the experiment. In Chapter 6.3.1 I compare the normative predictions derived from the equilibrium analysis with the actual decisions made by subjects in the experiment. Chapter 6.3.2 provides reasons for subjects' deviations from the normative equilibrium predictions. In Chapter 6.3.3 I show results for the propositions and the hypotheses. Chapter 6.4 summarizes the findings, highlights my contribution to the literature and sketches recommendations for future research.

Finally, Chapter 7 synthesizes the entire dissertation, including research questions as well as results and recommendations. The chapter is structured in three parts. In Chapter 7.1 I wrap up the conclusions from Chapters 2 to 4 which mostly dealt with the issue of optimal patent protection. Chapter 7.2 is concerned with patent races only, corresponding to Chapters 5 and 6. In the last subchapter, 7.3, I provide recommendations and point out to potential future developments concerning both patents and patent races.

CHAPTER 2

Economic Models of Optimal Patent Regulation

"Economic theory makes predictions about the effects of policy parameters that are sometimes quite sensitive to model assumptions, and it is often difficult to connect specific changes in patent rules and practices to the theoretical constructs."

Jaffe (2000)

"We have shown that the patent breadth-length optimal mix depends in a subtle way (involving second derivatives) on the relationship between social welfare and post-innovation profits, on the one hand, and the breadth of the patent, on the other hand. And economic theory places no restriction on the concavity of these functions. Thus it should not be surprising that different models and examples yield seemingly contradictory conclusions. But what is the economic intuition underlying these diverse results?"

Denicolò (1996)

Patent policymakers must play a delicate game of balancing between two extreme conditions. At one end there is too strong patent protection, which brings overly high rewards for the innovator through a monopoly position but punishes society who has to incur the deadweight loss. At the other end there is too low patent protection, which destroys the incentives to innovate altogether and therefore benefits no one. In this chapter I first describe instruments of patent regulation. Second, I examine the literature on optimal patent policy and the effects of this policy on welfare and innovation.

2.1. Instruments of Patent Regulation

A patent granting authority has many instruments at its disposal, through which it is able to regulate the appropriability of profits stemming from innovations. According to Judd et al. (2012), national patent systems use a set of tools such as “filing requirements, duration and scope of a patent, and renewal fees”. The literature on patent regulation speaks of these tools in terms of patent breadth, height and length. I describe them in the following subchapters. At the end of Chapter 2.1. I provide a table of definitions for these concepts, based on the literature.

2.1.1. Patent Breadth

An important element which determines the value of a patent is the patent's breadth, a concept defined quite ambiguously in the literature. In his article from 1972, Nordhaus firstly introduces the concept of patent breadth and describes it in terms of the appropriability of cost reductions induced by a process innovation. Nordhaus assumes that an innovator patents a technology which reduces production costs from a high cost C_H to a lower cost C_L. If B, the breadth of the patent, is zero, then competitors can imitate the process innovation perfectly easily and also reduce their costs by C_H - C_L. If B = 1, there is perfect patent protection and competitors may not reduce their costs at all, being forced to pay the high cost[1]. Nordhaus's definition of patent breadth is therefore related to the proportion of cost savings generated by the patented process innovation that is freely available to non-innovating competitors.

[1] Nordhaus (1972) uses a different notation from the one in this dissertation.

A rather different definition can be found in Klemperer (1990), who sees patent breadth as the “region of (differentiated) product space covered” (p. 116), meaning that “breadth measures how different competitors' products must be in order not to infringe the patent”. Following Klemperer's definition, a broader (or wider[2]) patent means that a non-infringing competing product needs to contain more differentiated characteristics compared to the patented product. Interestingly, Klemperer notes that in 1793 patent breadth was virtually equal to zero, and competing inventors could “file almost identical applications” to get patents of their own. Patent breadth came into existence through US acts from 1836 and later (idem). Gilbert and Shapiro (1990) define breadth very generally as any aspect of patent policy that affects the flow of profits to the patent holder. Matutes et al. (1996) define patent breadth as “the number of applications of the basic innovation which the patent holder is allowed to use in exclusivity”[3]. Van Zeebroeck et al. (2009) regard the number of claims per patent as a measure of patent breadth, while Lanjouw and Schankerman (1997) relate breadth to the probability that a patent is litigated. Finally, van Dijk (1996) also has his own definition of patent breadth (given in the next chapter), but he puts it in relation to another regulating instrument: patent height.

2.1.2. Patent Height

The height of patent protection is defined as “the stringency of novelty requirements that patent offices use in judging patentability” (van Dijk 1996). Patent laws demand that inventions can only be patented if they are non-obvious and include an inventive step, but the task of defining the exact boundaries of patentability falls into the responsibility of patent offices (idem). Van Dijk (1996) writes that it is fairly easy to define novelty requirements for fundamental inventions, but the judgment is more difficult in case of improvements of existing products, because it is hard to decide whether an improvement is obvious or not and what degree of inventiveness is required for patenting. However, patent history shows that most patents are conferred to product improvements (Baker, 1976) and this makes patent height an important, albeit delicate dimension of patent protection.

---

[2] Klemperer (1990) uses the terms "patent breadth" and "patent width" interchangeably

[3] Matutes et al. (1996) also use the term "patent scope"

Van Dijk (1996) defines the difference between patent breadth and patent height, which he considers fundamental. Patent breadth forms restrictions on the *imitation* of patented products, in that it sets a “maximum number of product characteristics that are allowed to be imitated” (van Dijk 1996, p. 152). Patent height sets conditions on product *improvements* (and not imitations) “by requiring a minimum number of new product characteristics” (idem). Van Dijk distinguishes between imitations and improvements from three points of view: costs, legal aspects and demand aspects. In the following, I closely follow van Dijk’s argumentation. From the cost side, imitations require some R&D investments in order to obtain learning effects which make production worthwhile; however, improvements necessitate even higher R&D costs, because, in order to make product characteristics more efficient or come up with new characteristics, additional knowledge needs to be acquired. From the legal side, “a sufficient improvement is in principle patentable, while a sufficient imitation is not” (idem). From the demand side point of view, imitations can be regarded as horizontal product differentiations: some consumers will prefer the imitation to the original if prices are equal; in other words, the market is split between originals and imitations. Improvements can be viewed as vertical differentiations: at the same price, all consumers should prefer the improvement to the original product, as the former brings buyers a higher utility. Therefore, from the demand side being an imitator does not secure any market advantages, while being an improver promises higher payoffs (idem).

Scotchmer and Green (1990) study the effect of novelty requirements (or patent height) on research and patenting decisions. They develop a model in which firms decide to enter or exit a patent race and examine the effects of a high patent, corresponding to strong novelty requirements, *vis-à-vis* a low patent height in the form of weak novelty requirements. Some of the most relevant observations made by Scotchmer and Green are the following: strong novelty requirements serve the purpose of protecting innovators’ profits by deterring competitors from making small product improvements, as these would infringe the patent; in this regard, high patents keep alive firms’ initial incentives to innovate. By contrast, weak novelty requirements support the social goal of information disclosure, which brings significant advantages on its own: since more technical knowledge is available to firms in an industry, these firms are more likely to innovate and thus the speed of innovation is overall enhanced. However, the authors point out to a drawback of disclosure, in that it "confers a positive externality on a firm's competitors, which the firm might want to avoid" (idem). Indeed, under weak novelty requirements a strategic player might refuse to patent or market small improvements, in order to block other players from using the knowledge made available. Due to this strategic component, low patents may not fulfill their desired goal of information disclosure.

Van Dijk (1996) develops a standard vertical differentiation model to analyze the effects of patent height. Here, two firms compete in product improvements and only one firm holds the patent on the basic invention. Both firms choose their innovations, but the non patent holder has a limited choice due to the novelty requirements implied by the patent height. Van Dijk's analysis is based on a non-cooperative duopoly competition consisting of three stages: an entry stage (in which firms decide to enter or not), an improvement stage (where firms decide on the level of improvement) and a price stage. The author notes that this order of the stages reflects increasing decision flexibility for the firms. After characterizing the problem mathematically and solving for equilibrium, van Dijk comes up with three sets of results. First, low patent protection does not influence the market outcome that would have been obtained without a patent, so low patent height plays no role in the competition. Second, high protection guarantees the patent holder a monopoly position at the detriment of the non patent holder. The third and most interesting case can be found for an intermediate patent height. Here, the novelty requirements imposed by the patent office "serve as a commitment for the non patent holder to choose the more profitable improvement" (idem). This leads to a "patent trap" for the patent holder, who loses access to the more profitable equilibrium. Van Dijk points out that his model is highly stylized, for which reason "applications and predictions must be taken with great care". Still, he finds empirical evidence that many firms are deterred from patenting by the application costs in the case of low patent height, which, as shown above, does not influence competition. He also finds empirical support for the "patent trap" in the intermediate case by comparing nations with different patent heights. Van Dijk's most important result here is that different patent heights can be used as instruments to encourage basic research versus applied research and development. He notes that in Japan, where patent protection is rather low because claims can be patented separately, a relative advantage in applied research and product improvements can be observed. Differently, in Germany and the US, where patents are high, a comparative advantage in basic research exists (idem). To my mind, this implies that patent height is more likely to lead to radical innovations by encouraging basic research.

2.1.3. Patent Length

From the three types of tools I am discussing, patent length is the most comprehensible, can be described quite straight-forwardly and is also the preferred instrument of regulation by social planners due to the easiness to define it and modify it. Patent length represents the lifetime of the patent, the length of time for which the patent is valid and generates benefits for its holder. The formal concept of patent length, also called *de jure* life (Scotchmer and Green 1990) or statutory life of the patent (O'Donoghue et al. 1998), must be clearly delimited from the concept of effective life of a patent. The latter is defined as "the expected time until a patented product is replaced in the market" (idem). There is a very large difference between the two concepts. While the *de jure* patent length is generally 20 years, the effective lifetime of a patent is much shorter: Mansfield (1984) brings survey evidence showing that, in certain industries, 60 % of patented innovations were imitated in no more than four years after their introduction. Especially in cumulative innovation settings, where technologies are developed quickly and original innovations give way fast to subsequent products, imitations may appear even faster. Levin et al. (1987) report that almost all patented products are duplicated in less than five years.

Nordhaus (1972) builds a model of optimal life for patent protection restricted to small process inventions. The model is based on several assumptions, from which I selected the most important: i) perfectly inelastic supply of inventors; ii) complete patent protection; iii) competitive product and factor markets. The only potentially critical assumption here is i). However, since the number of existing firms which invest in R&D does not change dramatically over time, I can consider this assumption to be quite realistic. Using well-chosen estimates for the parameters in his model (among others, price elasticity of demand, level of inventive inputs, percentage of cost reduction, private and social discount rate), Nordhaus reaches the following conclusions:

- for a natural or regulated monopoly, patents and the length of patent life are irrelevant to the level of welfare, and the inventing firm can capture its payoffs irrespective of the existence of a patent;
- in the case of several large firms and significant entry barriers, the general welfare problem cannot be solved through patent regulation;
- for relatively easy (trivial) inventions, the life of the patent is too long and should be shortened, or patents should not be conferred in the first place;
- for small innovations (which lead to cost reductions of less than 5%), the monopoly losses induced by the patent system are rather small (less than 20% of the gains from innovation);
- for drastic innovations (which bring significant cost reductions), a firm receiving a patent will lower its price by a portion of the cost saving for competitive reasons: in order to increase its market share. If the firm successfully obtains a larger market share, the resulting output will be produced at lower cost, which will then increase the average efficiency in the industry (idem). The gain to the economy during the lifetime of the patent will exceed the private gain of the firm, and some of the benefits will be immediately transferred to consumers (idem). Thus, for drastic innovations Nordhaus recommends that the optimal patent life should be slightly increased.

The definitions of patent regulating instruments encountered in this chapter are summarized in Table 2.1.

2.2. Optimal Patent and Patent Race Regulation

“Optimality” in the context of patent regulation may refer to a variety of concepts. The corresponding literature is highly heterogeneous; it ranges from topics such as the effects of patents on welfare to the impact of patent systems on innovation, to patent policy in patent races and to further ramifications, e.g. the effects of intellectual property on economic growth and inequality (Chu and Peng, 2011). In a sense, the literature on optimal patent regulation presents similar drawbacks to the patent race literature from Chapter 5: it is highly complex and based on different models under

Table 2.1. Definitions of regulating instruments

Concept	Author	Year	Definition
Patent Breadth	Nordhaus	1972	Appropriability of cost reduction induced by a process innovation
	Klemperer	1990	Region of differentiated product space covered
	Gilbert and Shapiro	1990	Any aspect of patent policy that affects the flow of profits to the patent holder
	Matutes et al.	1996	Number of applications of the basic innovation which the patent holder is allowed to use in exclusivity
	Van Dijk	1996	Restrictions on the imitation of patented products
	Lanjouw and Schankerman	1997	Probability of patent litigation
	Van Zeebroeck et al.	2009	Number of claims per patent
Patent Height	Scotchmer and Green	1990	A high patent means strong novelty requirements
	Van Dijk	1996	Restrictions on improvements of patented products
Patent Length	Mansfield	1984	The effective lifetime of a patent is much shorter than the statutory life of a patent
	Scotchmer and Green	1990	*De jure* life of the patent, typically 20 years
	O'Donoghue et al.	1998	Statutory life of the patent

miscellaneous, mostly incompatible assumptions. Kamien and Schwartz (1974) recognized early in the patent literature that establishing an optimal patent structure will depend on the interplay of a large number of influencing factors. In this chapter I have identified three main streams of the literature, which are discussed below.

2.2.1. Welfare Tradeoffs Between Regulating Instruments

Much of the literature considers patent length in its relation with patent breadth. Nordhaus (1972) finds that patent "life and breadth go hand in hand": if the patent breadth should be reduced because it causes undesirably strong protection, the patent length can be increased for compensation. In other words, there is a tradeoff between patent breadth and length. Similar recommendations are given by Gilbert and Shapiro (1990), who state that "whenever patent breadth is increasingly costly in terms of deadweight loss" (p. 111), infinitely-lived patents are optimal, but with "minimum market power necessary to attain the required reward level" (p. 107). What Nordhaus as well as Gilbert and Shapiro refer to is that broad patents may give an innovator a strong and wide monopoly position. The innovator can use this position to block competitors from developing products around the broad patent and sell the product at high price, with severe consequences on consumer surplus and a large deadweight loss. A regulator seeking to maximize social surplus should therefore reduce patent breadth to a minimum and largely increase patent length. However, Gilbert and Shapiro do recognize an important limitation of their paper: their conclusion only works in a stationary environment; in practice, inventions build on one another and thus a long patent might delay further innovation.

The study by Klemperer (1990) represents an important extension to the one by Gilbert and Shapiro (1990), by taking into account the role of consumers' substitution cost. Klemperer rightfully shows that the optimal decision on patent breadth depends on whether demand is more elastic in price than in substitution cost. If demand is more elastic in price, then Gilbert and Shapiro's conclusions are correct: to ensure demand can be satisfied, price must be kept low and therefore a minimal patent breadth combined with maximal patent length should be enforced (so as to avoid any monopoly or deadweight loss). If, however, demand is more elastic in substitution cost than in price, the converse is true and a broad but very short-lived patent is preferred. In other words, Gilbert and Shapiro's model is correct in the context of a homogenous product, while Klemperer's is accurate for differentiated products. Klemperer gives examples of products for each of the two situations. A product with high price elasticity of demand could be a computer program where different competing versions exist, some of which may be more expensive but easier to learn and use than others. If consumers are more sensitive to price than to the effort of learning, the patent for the computer program should be long-lived and of small breadth. A product with high elasticity of demand in substitution cost could be a drug that cures a serious disease but where other versions of the drug produce side effects. To my mind, this example is not well chosen, because it implies setting a large patent breadth for a product whose competing alternatives might produce negative side effects to people. In such a case, the monopolist patent holder could set a very high price for the product, forcing many consumers to buy alternatives that would damage their health. Instead, I would like to give here my own example of an appropriate product: a luxury watch with Internet connectivity where consumers have high substitution cost due to brand loyalty should be granted a broad and short-lived patent. Klemperer himself writes in his concluding chapter that his recommendations are based on a simple stylized model and drawing policy conclusions should only be done under extreme caution.

Both papers by Gilbert and Shapiro (1990) and by Klemperer (1990) are limited by their stationary point of view. In the remainder of this subchapter, I discuss the paper by O'Donoghue et al. (1998), which is based on the more realistic assumption that innovation is sequential and firms continually improve on each other's products. Indeed, Hopenhayn et al. (2006) write: "A central feature of innovative activity is that research is cumulative. […] New innovations build on the knowledge embodied in previous innovations". Scotchmer (2004) provides many examples for the importance of sequential innovations, may it be for software, which is based on previous generations, for hardware, which builds on earlier less developed products, or for biotechnology, where each new crop is used as a base for subsequent crops.

O'Donoghue et al. (1998) discuss patent policy in such a sequential innovation setting and note that "a patent can terminate either because it expires or because a non-infringing innovation displaces its product on the market". The time until any of these two happens is called the "effective life of a patent" (idem), and its difference to the statutory life was described in Chapter 2.1.3. In the paper by O'Donoghue et al., the effective patent life is shown to depend on patent breadth, because breadth sets conditions on which products may compete with, and ultimately supersede the patented product (independent of the statutory life). The authors distinguish between "lagging breadth", which protects against imitation from products of inferior quality to the patented product, and "leading breadth", which protects against new product improvements from superior alternatives (in van Dijk's 1996 terminology from Chapter 2.1.2, lagging breadth would be called patent breadth, while leading breadth would be patent height). O'Donoghue et al. show that lagging breadth would not generate enough stimulus to invest in R&D while, conversely, leading breadth could prolong the effective patent life and encourage R&D. They even go one step further into saying that "without some form of leading breadth, the effectiveness of the patent system to promote innovation is seriously impeded" (idem). Their main conclusions are that leading breadth is essential, as without it innovativeness may become "seriously suboptimal", and that a specified rate of innovation can be obtained via two distinct policies: i) a patent of infinite length but small leading breadth (height), or ii) a patent of infinite leading breadth (height) and short length. According to the authors, the difference between the two policies is that the first one would help to reduce R&D costs, while the second would reduce the costs of delayed innovation diffusion and alleviate market distortions.

In this subchapter, one could see that the choice of optimal patent regulation instruments is influenced by several factors: definition problems (most notably because patent offices will have difficulties to define and operate with concepts such as patent breadth and height or "leading" versus "lagging" breadth), by the characteristics of the innovation (e.g. stationary or sequential innovation, elasticity of demand with respect to price or to substitution costs), and especially by the welfare goal set by the regulator (e.g. reduce wasteful duplication of R&D efforts, boost the speed of innovation, or increase consumer surplus and decrease the deadweight loss). Thus, the optimality of regulating instruments will always depend on the entire context, any choice will generate some tradeoffs and a "one size fits all" instrument cannot be found.

2.2.2. Patent Protection and Innovation

In this subchapter I report on the most important results of the literature dealing with the effects of patents and patent systems on innovation. The study by Jaffe (2000) reviews some major changes that occurred in the US patent system between 1980 and 2000 and attempts to quantify the influence of these changes on technological innovation. Jaffe writes that the changes to patent policy and practice in the US have had the general effect of strengthening patent protection “institutionally, geographically and technologically”. He also notes that these policy changes coincided with an unprecedented surge in patent applications (which I describe extensively in Chapter 3). Since important modifications to the system of property rights occur very rarely, Jaffe declares them to be “an unusual set of natural experiments” which should help us better comprehend the effects of patent protection. Jaffe groups the changes in US patent policy into four categories and discusses them in detail in his paper. I briefly summarize them here, based on the information provided by the author: i) the creation of the Court of Appeals for the Federal Circuit, which examines previous patent decisions and increases the probability of success in court for patent holders; ii) the enhancement of “patenting and licensing privileges to inventors in universities and government laboratories who create commercially exploitable inventions”; iii) widening patent rights to new technology fields, especially software and gene research; iv) the agreements to expand patent protection and make it more compatible worldwide, which were formulated under the Uruguay Round of the General Agreement on Tariffs and Trade (GATT). Unfortunately, Jaffe’s paper only brings – in Jaffe’s own words – “disappointing results”. His main conclusion is that “despite the significance of the policy changes and the wide availability of detailed data relating to patenting, robust conclusions regarding the empirical consequences for technological innovation of changes in patent policy are few” (idem). Jaffe gives some reasons for the lack of solid conclusions: too many influencing factors on innovation change simultaneously, making it rather impossible to isolate the effects of policy changes; patents are only one of many elements which act upon innovative behavior; the predictions of economic theory are sensitive to the assumptions of different models of optimal policy; it is problematic to link the theoretical entities to the empirically observed phenomena.

The study by Sakakibara and Branstetter (2001) pursues a similar goal to that of Jaffe (2000). The former authors try to find evidence of changes in R&D spending or innovative output, as a response to the Japanese patent law reforms from 1988, which significantly strengthened patent protection. Sakakibara and Branstetter first describe the changes to the Japanese patent system: i) the reforms broadened patent scope by allowing multiple claims (independent or dependent on other claims) to be included in one patent. According to patent experts, after 1988 the scope of protection in Japan equaled or even exceeded that of the US and European patent systems (Okamoto et al., 1996); ii) patent term restoration for pharmaceuticals was extended by up to five years if delays in “the period necessary for drug safety and efficacy examinations” occurred (Sakakibara and Branstetter, 2001). According to the latter authors, this decision gave pharmaceuticals a prolongation of effective patent lifetimes. One effect of the patent reform was that the growth rate of patent applications decreased after 1988. This development seems natural, because the reform allowed more claims to be included in one patent and inventors must have taken advantage of this new option to save on application costs. While applications leveled off after the reform, a large increase in the number of granted patents was observed. The increase in grants was a result of the reformed multi-claim system, which made “the improvement of an invention over the existing technology easier to demonstrate” (idem). Another effect of the reform was a vigorous rise in the number of intellectual property lawsuits in Japan. Finally, Sakakibara and Branstetter built a dataset of 307 publicly traded Japanese manufacturing firms from different industries and performed an econometric analysis to establish the effect of the reform on R&D and therefore on innovation. Their results found no evidence of an increase in R&D spending or on innovative output that could be ascribed to the patent reform. This holds true for all industries taken into account, even for pharmaceuticals, which experienced particularly strong protection. Sakakibara and Branstetter’s findings corroborate the study by Jaffe (2000) as well as other results from the literature, which I summarize in the following.

Hall and Ziedonis (2001) studied patenting in the semiconductor industry and concluded that strengthened intellectual property rights did not induce additional innovation; instead, the stronger rights have led to a socially wasteful increase of defensive patent portfolios. Schankerman (1998) analyzes a dataset of all patent applications at the French patent office between 1969 and 1982 and shows that the private value of patent rights (computed as returns from holding the patent minus patenting costs) is very small for a large majority of patents (about 95%). He concludes that patent rights have value, but they are not the main source of returns from inventive efforts. Cohen et al. (2000) suggest that patents are relatively weak and imperfect instruments of appropriation of returns, and increases in protection might be insufficient to induce more innovation. They also point out that the patent system reforms encouraged the use of patents as legal weapons, as seen in the augmentation of intellectual property lawsuits. Lerner (2009) searches for effects of changes in patent law in several countries by looking at changes in the quality and quantity of foreign patents filed in the UK. He also finds no evidence of effects on innovation after reforms in patent systems. Encaoua et al. (2006) write that some countries have experienced a “weakening of the standard criteria for granting patents”, following a belief that “more patents are better”, but this tendency had negative effects on competition and on subsequent innovation. Last but not least, Hunt’s (2006) article develops a simple model that links firms’ R&D and patenting decisions. Hunt shows that if firms are sufficiently active in their R&D and patenting activities, incremental reductions in the cost of attaining patents would result in less R&D and less innovation. He therefore recommends raising patenting costs to induce more innovation. The main conclusion from this subchapter is that it is unclear whether patent protection may always enhance innovativeness; indeed, under some circumstances the opposite might be true.

2.2.3. Optimal Rules For Patent Races

Patent races and their features are extensively analyzed in Chapter 5. However, the issue of how to design optimal rules for a patent race given certain endogenous regulating instruments belongs to the current chapter, and I discuss it on the basis of three most important papers from the literature. Denicolò (1996) examines optimal patent breadth and length in the context of several firms that are racing for a patent. Denicolò first reviews the existing literature (most notably the papers by Gilbert and Shapiro (1990) and by Klemperer (1990), which I discussed in Chapter 2.2.1) and states that in the search for optimal patent regulation “almost anything could happen” and there is no regulating instrument which by itself can be optimal. Denicolò therefore develops a patent race model with several firms and many parameters: a firm’s R&D investments, a hazard function giving the probability of success dependent on R&D and time, the probability that a firm’s rivals innovate, an interest rate, a current profit of a firm, the value of the prize for the winner of the race and for its losers. Denicolò’s results are that the optimal mix between patent breadth and length will depend on whether social welfare and post-innovation profits as functions of patent breadth are concave or convex. He notes that, since economic theory puts "no restriction on the concavity of these functions", the analysis of different models from the literature will lead to contradictory results. He suggests an answer to these contradictions. Generally, a reduction in patent breadth will lead to intensified competition in the product market after the patent race is finished; this happens because a limited breadth allows more competitors to innovate around the patent. But, as Denicolò puts it, more competition is not always desirable, because it may imply social costs such as inefficient production or the duplication of entry costs. Thus, if a less broad patent generates socially costly competition, it is better to award patents with maximal breadth and minimal length. Conversely, a reduction in patent breadth is optimal if more competition increases social welfare without reducing too much the incentive of firms to engage in innovation and participate in the patent race (idem).

The paper by Fershtman and Markovich (2010) presents a two-firm asymmetric-ability multistage R&D race model depending on a weak or strong regime of patent protection. In Fershtman and Markovich's setting, one firm has "a technological advantage in the early stages of the race", the other firm in the later stages (in the experiment from Chapter 6, I also analyze an asymmetric setting in which one firm has a time advantage). Their model has the strength that it accounts for the accumulation of knowledge during the race. The authors try to accomplish a difficult endeavor – they examine "the effect of patents, imitations and licensing arrangements on the speed of innovation, firm value and consumers' surplus" (idem). They set up two extreme regimes, one of strong patent protection at the end of the race, with a "winner-take-all" feature, and another one of weak protection under which any discovery can be completely imitated. They discuss which of these regimes performs better, depending on three variables: i) the final prize, defined as the market value of the patented product minus the cost of innovation; ii) the intensity of the duopolistic competition, measured as μ, the ratio between duopolistic profits and the profits that would be obtained under monopoly, where a stronger duopolistic competition is reflected in a lower μ; iii) the degree of asymmetry between the firms' R&D abilities. Fershtman and Markovich provide a numerical analysis of such R&D races and solve it for equilibrium using a wide range of parameter values. Their main finding is that in a dynamic R&D race with asymmetric abilities, a weak patent protection in the form of perfect imitation may induce higher consumer surplus and higher value for firms than strong patent protection. However, this result only holds under certain settings, which are specified by the authors: first, the result is valid if the two firms' "asymmetry in R&D abilities is sufficiently large", but not if the firms "have identical or similar R&D abilities" (idem). Second, if the market for the patented product is small when compared to the cost of innovating, or if the duopolistic competition is strong (leading to lower market benefits for the participating firms), the weak patent protection regime fails to induce enough stimulus for innovation and therefore, the conventional reasoning for strong patent protection holds. Fershtman and Markovich also give explanations for their results, which are the result of two conflicting tendencies. On one hand, an R&D race under weak patent protection (imitation) "always ends up with a duopoly - implying a lower prize at the end of the race and thus lower incentives to invest" (idem). On the other hand, the R&D process itself may work more efficiently under the imitation regime (a reason for this is given by Bessen and Maskin (2009) who discuss sequential innovation; I reported on their paper in Chapter 1.1.4). But when firms' R&D abilities are identical and imitation is possible, a free rider problem will emerge: a firm will wait for others to innovate and then copy the innovator's product with very little effort. According to Fersthman and Markovich, the free riding phenomenon "reduces the firms' incentives to invest in R&D which implies a slow pace of innovation and consequently low consumers' surplus and a low value for firms". For this reason, the weak protection or imitation regime will only work if firms are sufficiently dissimilar in R&D abilities. Fershtman and Markovich also analyze the effects of licensing, which is beyond the scope of this dissertation. Finally, the authors write that "the effect of different patent regimes on the race depends mainly on the detailed characteristics of the race. Clearly, one cannot find one regime which dominates all the others for all possible R&D races. […] A general claim about the appropriate optimal patent policy is beyond the scope of this paper".

The third paper of high relevance for this chapter is the one by Judd et al. (2012), who derive optimal rules for patent races. Their paper lies at the intersection of the literature on patent regulation and the literature on patent races. I discuss it extensively in Chapter 5.1.4, but I only give a brief description here. According to Judd et al., a patent regulator can influence the optimality of a patent race by using two regulating instruments of regulation: i) the length of the race, defined as the "minimal accomplishment" required from the innovating firm in order to be granted a patent (in the terms of the current chapter, I would label this as patent height); ii) the size of the prize, equal to the benefits for the innovating firm from receiving of a patent. However, Judd et al. point out that any regulation of these two instruments confronts the patent granting authority with a series of very difficult trade-offs, and any attempt by the regulator to make an optimal decision eventually results in a vicious circle. After solving a complex dynamic multistage innovation race model with a variety of parameters, Judd et al. draw conclusions on optimal regulation, which is found to depend on various specific objectives of the regulator. The authors propose two alternatives to the regulation of patent races: i) to finance the prize through taxation and give the prize to the inventor, but make sure the innovation is produced on a competitive market; ii) research tournaments, where "contestants compete to find the innovation with the highest value to the sponsor" (idem).

This chapter gave detailed definitions of the instruments for patent regulation: patent breadth, height and length. Patent length is the preferred instrument of regulators because it is easy to define and modify. Nordhaus (1972) shows that strong patent protection in the form of a long-lived patent should only be chosen for radical innovations and in virtually no other case. However, even Nordhaus's conclusion may be refuted if one considers sequential innovation on the lines of Bessen and Maskin (2009) or Hopenhayn et al. (2006). Besides, as pointed out by Jaffe (2000) as well as other researchers, patent breadth and height are abstract concepts that cannot easily be linked to patent practices. Therefore, patent offices may only find limited use for many results of the literature on regulation. In Chapter 2.2 I searched the literature for optimal regulating instruments and could not find a "one-size-fits-all" solution. Each paper suggests answers that will depend on many market specificities as well as varying goals of regulators and firms. Overall, I could not find a clear link between the strength of patent protection and innovativeness.

CHAPTER 3

The Dangerous Development of Patent Applications in Europe

"Outrageously large applications with high number of claims are often filed as smokescreens, and contribute to polluting the collections of prior art, and their uncontrolled growth in voluminosity therefore contributes to a deterioration of the legal certainty for all users of the patent system."

Van Zeebroeck et al. (2008)

"The size and length of patent applications at the EPO has drastically increased over the past 20 years, raising serious questions as to the ability of patent offices to manage their workload while upholding high standards of quality in the granting process."

Van Zeebroeck et al. (2009)

"Is the increase in the number of claims actually revealing a better and more systematic use of fallback positions in patent drafting and number of pages a more thorough disclosure of inventions for which protection is being sought?"

Archontopoulos et al. (2007)

Since the 1980s, the number and size of patent applications and grants has been growing continuously worldwide, resulting in a dramatic increase in the workload of patent offices and, most importantly, raising "serious concerns over patent quality standards" (van Zeebroeck et al., 2009). A good example for this growth can be observed at the European Patent Office (EPO). The Office granted its first patent in the year 1978 and celebrated its 200,000th grant only 14 years later in 1992, recording an average of 14,000 patent grants per year (Wissensallmende, 2009, based on EPO data). This figure is nonetheless small compared to the 40,000 patent grants recorded at the EPO in the year 1997, and the 52,000 grants from 2007 (idem).

Even though a similar development could be observed in the US and around the world, this chapter discusses the evolution of patent voluminosity in Europe. After the *status quo* is described, the chapter provides explanations for the patent inflation, shows its effects on the system and suggests possible solutions.

3.1. How Patent Applications Exploded

Patent size, or voluminosity, has two dimensions: its total number of claims and its total number of pages. Both dimensions have radically surged at the EPO during the period 1980-2004. Moreover, the patent size increase was accompanied by a boom in the total number of patent applications, and these combined factors have put tremendous pressure on patent systems (see for example Archontopoulos et al., 2007).

3.1.1. Theory of Patent Claims and Pages

The number of claims is seen in the literature as a reflection of patent breadth – the scope of protection – because more subject matter is included with more claims (see Tong and Frame, 1994 or Lanjouw and Schankerman, 1999). This view is generally accepted, even though there are counterexamples related to the wording of the claims: using the word "rodent" would take one patent claim; instead, naming three different rodents would lead to three claims, even though the scope of protection is in this case much smaller (van Zeebroeck et al., 2009). Nevertheless, more claims "denote a more detailed definition of the protected area, adding precision: instead of giving a generic term which could be somewhat vague, the applicant will list extensively and individually all potential components of the subject matter" (idem). The rigorous descriptions included in more claims also provide patent holders with increased legal power in case the patent is litigated or contested. In such cases, patentees must defend their property rights against the objections of a legal examiner. They achieve this by "having a series of claims partly overlapping, partly fitted into each other" (van Zeebroeck et al., 2009) or, differently put, making claims "serve in different contexts" (Archontopoulos et al., 2007) and thereby maximizing the "chances for the main claim" (idem) to be considered valid. A different interpretation of claims is provided by Stevnsborg and van Pottelsberghe de la Potterie (2007), who regard the number of claims as a "strategic choice": the applicant may choose to use more claims in order to hide the real invention "in the middle of many non-inventions" and use "vagueness as a weapon" (idem). These authors explain that there are two possible reasons for such a strategic choice: either the applicant wants to mislead patent examiners and especially competitors (in order to avoid relevant knowledge from being publicly used and further developed), or the applicant simply does not know what the relevant invention is when she files for a patent. An example for the latter case, illustrated by the authors and reproduced slightly differently here, could be a developer of chemicals who performs research and discovers a class of molecules, out of which only some molecules have a certain property (for instance that they lower cholesterol levels in humans). In case the researcher does not know exactly which molecule has the property, she will file claims for all molecules of the class, in order to be sure that the most valuable discovery is included in the scope of patent protection. In my view, the strategic inclination of patentees signaled by the latter authors may also be one of the factors contributing to the upturn in patent voluminosity.

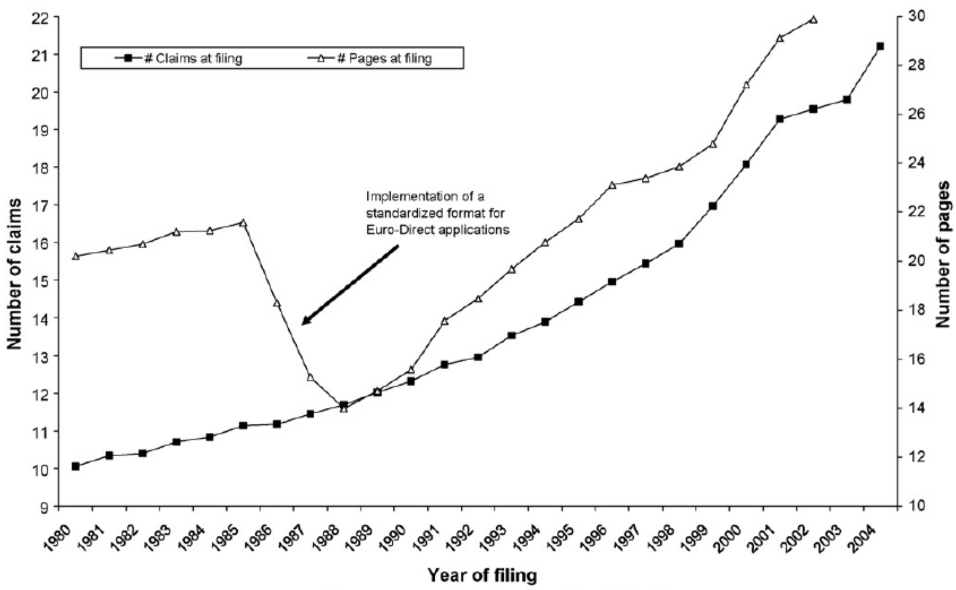
The number of pages in a patent application is considered to express "the extent of the disclosure of the invention", because more pages typically reflect more in-depth descriptions and more comprehensive drawings (van Zeebroeck et al., 2009). The number of pages partly results from the number and length of the claims: Archontopoulos et al. (2007) find a correlation of 0.37 (significant at a 5% probability threshold) between the number of claims and the number of pages at filing. Even though the number of pages may seem to be a straightforward measure of the size of a patent application, there are some limits to its unbiasedness – the number of pages is influenced by the format of the document (font size, margins, paragraph or line spacing), by its language and by the presence of illustrations (Archontopoulos et al., 2007). In order to alleviate some of these problems, the EPO has applied a standard format to Euro-Direct applications since the mid-1980s. Besides, the formatting issue disappears when one considers granted patents, because these are published by the EPO in a specific and very compact format called a "type-set format" (idem). The number of illustrations in applications is contained in the EPO's databases and therefore could be controlled for. Moreover, applications have to be provided in one of the three official languages: English, French or German, which somewhat restricts the language issue (idem). Some language heterogeneity issue still remains, but it is not severe – Archontopoulos et al. (2007) point out that the 2002 edition of the European Patent Convention comprises 73,629 words in German, 84,583 in English and 86,353 words in French; however, the number of characters in the three languages was less dissimilar, resulting in a comparable number of pages.

3.1.2. Dynamics of Patent Growth and the Role of R&D

Even though the phenomenon of patent voluminosity and complexity has become very stringent after the 1980s, it is not entirely new. Smith (2003) writes that as early as 1933 the US Patent and Trademark Office Society (USPTO) was already on the search for methods to eliminate multiple claims from applications and to design application fees depending on the number of claims. Duncan (1965) notes that the problem was repeated three decades later: patent offices were again confronted with increasing complexity, which led to delays in the handling of applications. As a result, patent offices employed more examiners, introduced automatic searching and modified procedures to raise filing and renewal fees (idem). In the 1990s and 2000s, applications of more than a thousand pages, which once could not be imagined, started to be filed quite often at the EPO and other patent offices worldwide (van Zeebroeck et al., 2009). Several applications even reached 100,000 pages or up to 20,000 claims (idem). The growth of patent size became so extreme that the term "mega-application" was introduced, "often in relation to applications filed together with biological sequence listings" (Archontopoulos et al., 2007). In one case, the EPO received an application containing 283 priorities and 80,259 sequences, totaling in around 600,000 pages (idem)!

Figure 3.1 shows the evolution of claims and pages between 1980 and 2004. The average number of claims per patent application more than doubled in this period, from 10 claims in 1980 to 21 claims in 2004. From 1985, one can observe a temporary reduction in the average number of pages, as a result of the implementation of a standardized format for applications at the EPO. In 1995, ten years after the standard format was introduced, the number of pages picked up from its 1985 level and continued to grow ever since. Similarly to the number of claims, the number of pages nearly doubled over the entire period, increasing from around 15 pages per application in 1980 to 30 pages in 2002. An essential piece of information not included in Figure 3.1 is that, additional to the increase in claims and pages, the total number of patent filings also grew – it doubled during the period 1998-2008 (van Zeebroeck et al., 2009), leading to a truly exponential growth trend.

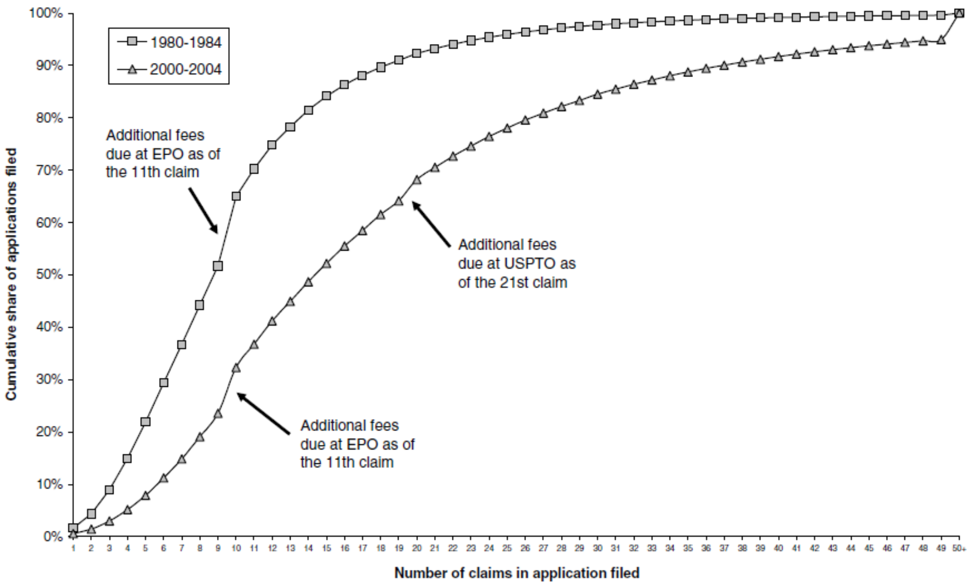
Figure 3.1. Average number of claims and pages in EPO applications between 1980-2004



Source: van Zeebroeck et al. (2009), based on EPO data

Figure 3.2 provides a different way of looking at the evolution of patent voluminosity. Here, two cohorts of EPO applications filed in the periods 1980-1984 and 2000-2004 are analyzed. The figure shows the evolution of the cumulative share of applications filed in the two periods and their corresponding number of claims. The distributions of claims contain two distinct modes, one at 10 claims and another one at 20 claims. These modes reflect the additional claims fees requested at the EPO (starting with the 11th claim) and at the USPTO (from the 21st claim). A large share of EPO applications is concentrated between the 9th and the 10th claim, showing that applicants tried to avoid the additional fees by restricting their protection to exactly 10 claims (similarly, many USPTO applications were restricted to a number of 20). The most important and striking result from Figure 3.2 is the shift of the distribution towards the right side and a significant lengthening of the distribution's tail between the two cohorts. For example, in the period 1980-1984, 90% of applications contained 19 claims or less; twenty years later, between 2000-2004, 90% of applications contained 37 claims or less, a virtual doubling in the number of claims. Furthermore, between 1980-1984 there were almost no applications with more than 50 claims; between

Figure 3.2. Cumulative frequency distribution of the number of claims in EPO applications



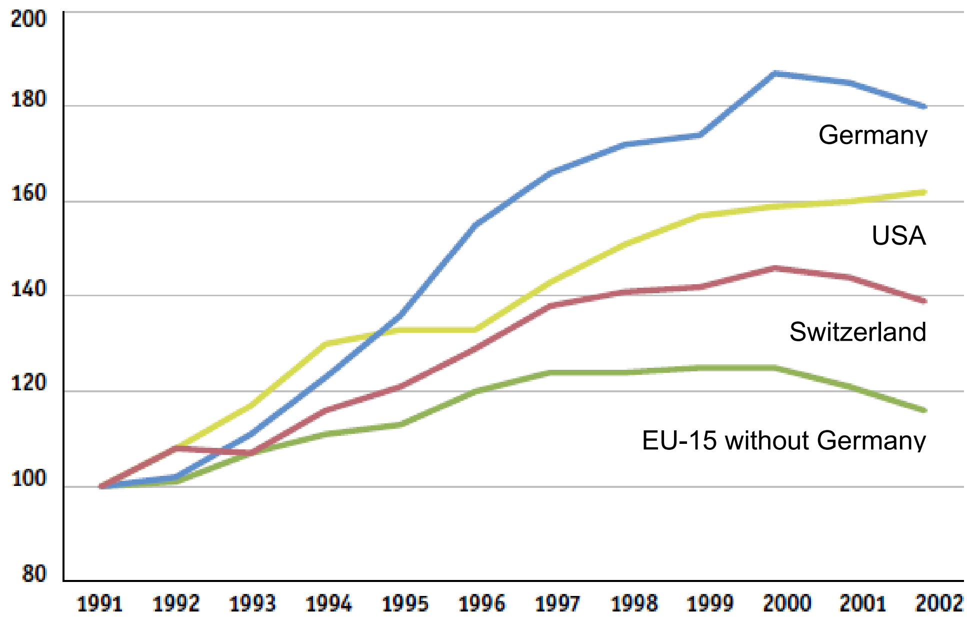
Source: van Zeebroeck et al. (2008), based on EPO data

2000-2004 there were about 5% of applications with more than 50 claims. Both the right shift and the lengthening of the distribution of claims are clear signs of the surge in patent voluminosity.

Since the main purpose of patents is to reflect inventive activities (as well as offer inventors protection from imitation), some important questions arise. Has the explosion of patent applications been accompanied by a corresponding increase in R&D activity? Or does it rather reflect a higher tendency of firms to patent more? Since many studies indicate that R&D activities are a central input for inventions, and thus for patent applications (see e.g. Licht and Zoz, 1998), one would expect the first interpretation to be the correct one. Unfortunately for society, the literature seems to provide support for the second interpretation. Cremers and Licht (2006) give evidence that the inflation-adjusted expenditures for R&D have barely increased in most countries in the 1990s. Instead, the ratio of patents to R&D expenditures visibly increased during that time. Figure 3.3 shows the development of the patent-R&D quotient between 1991 and 2002 for different regions of the world (it takes into account patent applications in the US). Following the logic of the invention production function developed by Griliches (1990), the patent-R&D quotient in any year is computed as the number of patent applications during that year divided by the R&D capital stock from the previous year. The patent-R&D quotient is normalized to a base index of 100 in the year 1991 for all regions of the world taken into account. Remarkable is the fact that, out of all regions, German companies seem to come up with much more patents from their R&D activities.

Kortum and Lerner (1999) add to the discussion on the patent-R&D quotient. They test the hypothesis whether a broad increase in research productivity underlies the growth in patenting by looking at the US research effort in the 1990s. Their data analysis shows no consistency with the formulated hypothesis. Instead, they find that the general increase in patenting was not accompanied by a surge in any particular research activity – research intensity leveled in the 1990s, while patenting continued to rise. The authors conclude that the boom in patenting must be due to improvements in the management of innovation of US firms, and not to R&D intensity.

Figure 3.3. Development of the patent-R&D quotient between 1991-2002



Source: Cremers and Licht (2006), based on OECD data.
Index for 1991 = 100

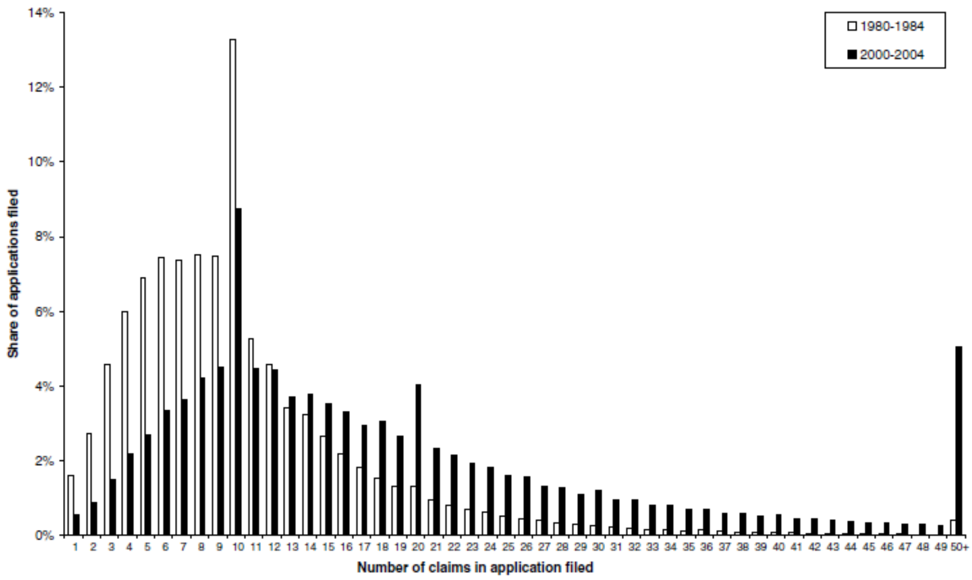
Chapter 3.1 has shown that the simultaneous boom in patent size and in the number of patent filings generated a tremendous workload for patent offices, which have to deal with a huge backlog of applications that have not been examined yet (Archontopoulos et al., 2007). More claims and more pages mean that patent offices have to assign more resources "to searching and examining the files, which induces an additional pressure on quality" (van Zeebroeck et al., 2009). The latter authors find it critical that the broader patent protection "may induce a higher cost to society, and possibly more uncertainty for competitors during the examination procedure and beyond" (idem). This uncertainty stems from the fact that, given a doubtful quality assessment for patent protection, competitors no longer know what the scope of protection really is and thus they cannot develop strategies to make sure they do not infringe the patent. Furthermore, we have seen in this chapter that firms' increasing tendency to patent was not coupled with changes in R&D activities. So, in the words of van Zeebroeck et al. (2009, p. 1007), one has to ask the question "whether the change is due to increasing greed on the part of patentees in claiming for more protection, or whether it is driven by structural and exogenous changes in patent systems, technologies and market conditions". The next chapter wishes to address this question.

3.2. Explanations for the Surge in Patent Voluminosity

3.2.1. Fees

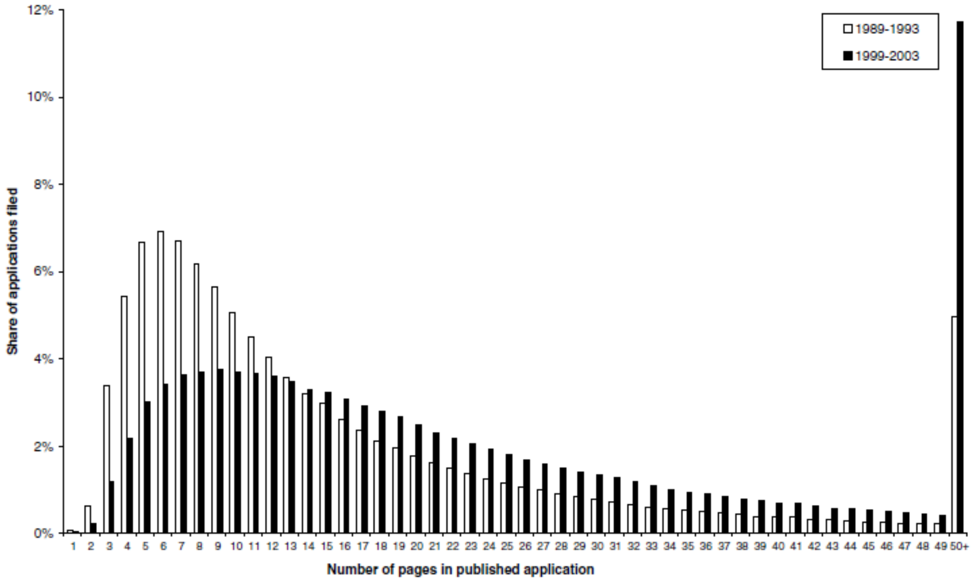
As one could observe in Figure 3.2 of the previous chapter, one factor with an impact on patent drafting and voluminosity is represented by administrative fees of patent offices. Figures 3.4 and 3.5 of this chapter give a more detailed overview of the distribution of both claims and pages for patent applications. The most remarkable result from the comparison of these two figures, also emphasized by Archontopoulos et al. (2007), is the different shape of the distribution for claims (tri-modal) versus pages (unimodal). The distribution of claims shows peaks at levels of 10, 20 and 50, while that of pages has only one mode at 50. The mode of 50 in both

Figure 3.4. Number of claims in EPO applications



Source: Archontopoulos et al. (2007), based on EPO data

Figure 3.5. Number of pages in EPO applications



Source: Archontopoulos et al. (2007), based on EPO data

distributions represents very large and mega-applications (which I discussed previously). But the first two modes in the distribution of claims are the ones revealing some interesting facts. Patent applicants appear to react to fee policies governing EPO and USPTO applications, which are based on claims, and not on pages. As shown in the previous chapter, the EPO charges extra fees starting with the 11th claim, while the USPTO does the same from the 21st claim. These two thresholds generated the first two modes in the claims' distribution.

In my view, this means that applicants had many applications consisting of 10 or 20 claims, in order to avoid additional fees. Furthermore, this most probably happens because patent offices introduced additional fees based on claims, and not on pages. In the words of Archontopoulos et al. (2007), patent applicants show some "price elasticity of the number of claims", but no price elasticity of the number of patent pages. Therefore, I think that a good solution for patent offices might perhaps be to change the current pricing scheme for claims (which at the moment is non-linear), and create progressive, monotone increasing fees that charge applicants for each additional claim. I will briefly discuss this as a possible solution as well as further solutions from the literature during Chapter 3.3.

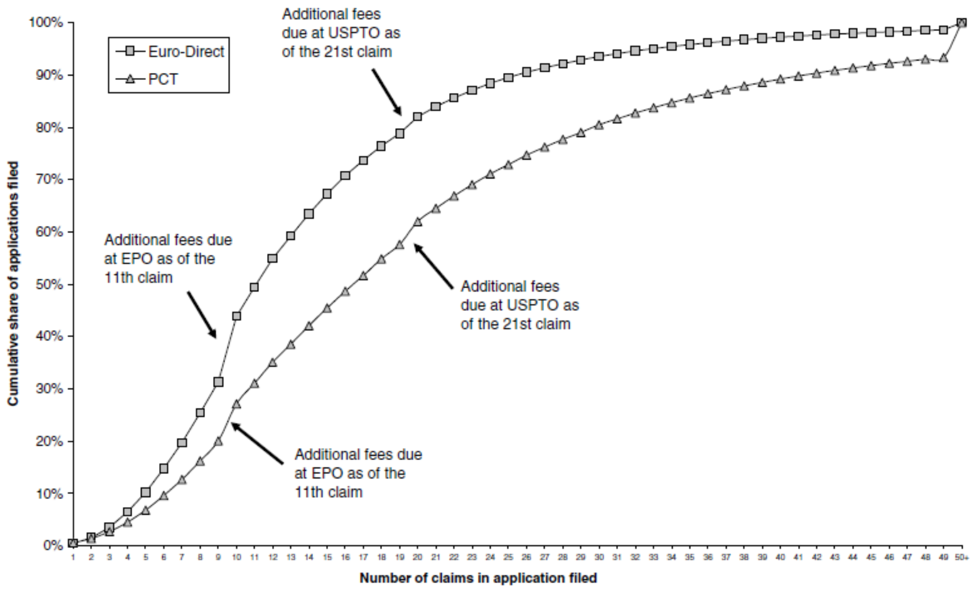
3.2.2. Internationalization

Archontopoulos et al. (2007) mention three possible routes that a patent application can follow before it gets filed at the EPO. It may start as: i) “a national priority filing subsequently transferred to the EPO”; ii) a direct filing at the EPO; iii) an international application through the Patent Cooperation Treaty (PCT).

Filing through a Patent Cooperation Treaty procedure may “impact the way a patent application is drafted” (van Zeebroeck et al., 2009). The latter authors explain the advantages of the PCT procedure. They note that through the PCT, which is proctored by the World Intellectual Property Organization, applicants may find an important method to reduce patenting costs. Van Zeebroeck et al. (2009) outline the PCT procedure: “Under this treaty, one can file an “International Patent Application”, which does not turn into some sort of international patent, but primarily acts as a vehicle to buy a period of time within which to proceed with national or regional (such as the EPO) patent applications. Instead of having only a 12-month time period within which an inventor must file foreign applications in order to claim priority, with the PCT, the inventor can gain an additional 18 months before having to incur the relatively large expenses of completing the applications at each of the designated offices around the world”. Van Zeebroeck et al. (2009) stress out the fact that this additional time “can be crucial to the exploitation of an invention” for a number of reasons, which I describe in the following: the time can be used to gather the significant financial resources necessary for a multi-country application; the inventor can search for agents who are willing to pay license fees to use the discovery, or for distribution partners, thus securing a portion of the future market; most importantly, the time can serve to reduce some of the uncertainty associated with the invention, by allowing the applicant to better assess the potential value of the discovery before a high cost is incurred. Last but not least, the PCT does not require additional fees for excess claims (idem), as is the case with direct applications at the EPO or USPTO (see for instance Figure 3.2). Therefore, the PCT makes it easier and cheaper for inventors to apply for a broader scope of protection. It was probably for all these reasons that the PCT option increased sharply from 15% of EPO applications in 1985 to a level of 50% in 2000, where it remained quite stable until the year 2005 and further (idem).

Figure 3.6 compares the cumulative frequency distribution of claims according to the filing route the applicant has taken: direct applications to the EPO (Euro-Direct) versus PCT applications at the EPO during the period 2000-2004. The distribution of Euro-PCT is shifted to the right and with a much longer tail than the Euro-Direct distribution, revealing that Euro-PCT applications contain significantly higher numbers of claims. To this, van Zeebroeck et al. (2008) add that PCT applications are "drafted according to US conditions" and are "more influenced by USPTO fee regimes". Van Zeebroeck et al. (2009) run an empirical model in which they test a set of hypotheses on the determinants of patent voluminosity, including: the internationalization of procedures through PCT, national practices, technological complexity and emerging sectors (which are explained in the next subchapters). From all these determinants, internationalization and national practices were found to have the highest explanatory power to account for the patent explosion. A patent filed under the PCT option was found to contain more than one claim and about 15 more pages than an average EPO application. Van Zeebroeck et al. (2009) conclude, just like the previous study, that PCT filings follow a "draft once,

Figure 3.6. Cumulative frequency distribution of the number of claims in Euro-PCT and Euro-Direct applications, 2000-2004



Source: Van Zeebroeck et al. (2008), based on EPO data

file everywhere" principle; they are "drafted with a US template then applied to all other patent offices". Most importantly, both studies mentioned above consider internationalization through the PCT route to be the main contributor to the surge in voluminosity.

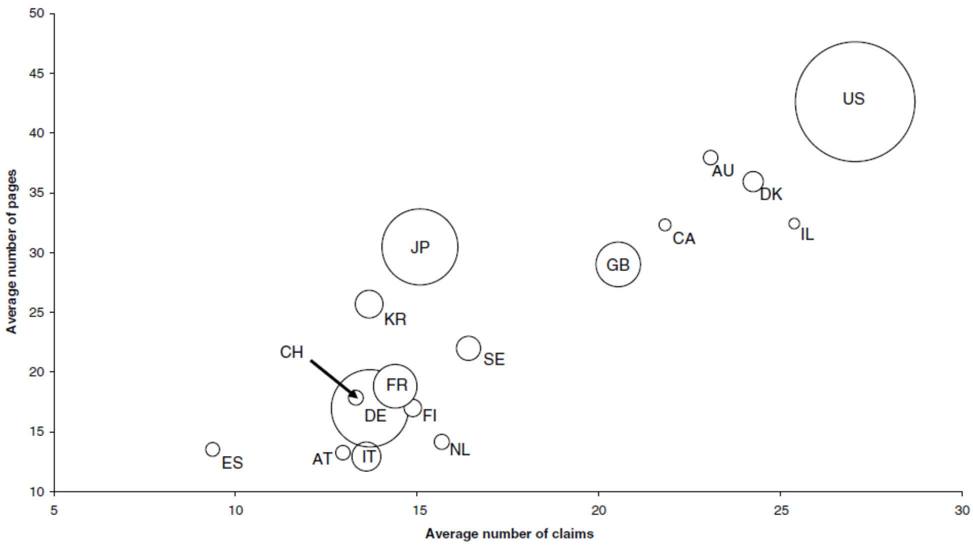
3.2.3. National Practices

Looking at differences in national practices also reveals large discrepancies in patent size. The largest differences can be found between the US and continental Europe. The literature attributes these differences to country-specific patent drafting styles, patent law as well as cultural factors. For example, van Zeebroeck et al. (2009) point out that commercial contracts are much larger in the US, showing differences in the legal system and culture that must have similarly affected US patent drafting practices.

Figure 3.7 displays the average number of claims and pages in EPO applications filed in 2002 according to their priority country. These geographical patterns show that two main groups of countries emerge. The first group, depicted in the upper-right portion of the figure, is mainly composed of Anglo-Saxon countries: the US, the UK, Australia and Canada.

This first group is characterized by applications with an above-average number of pages and claims. The second group, found in the lower-left part of the figure, is mostly made up of continental European countries such as Germany, France, Italy and Switzerland, which have smaller applications in terms of pages and claims. Van Zeebroeck et al. (2009) emphasize the fact that the main difference between these two groups of countries is their legal system, with common law governing the first group and civil law the second. It is the common law regime which leads to larger documents. Van Zeebroeck et al. (2009) also compared the size of patent applications for non-US applicants targeting the US market first with that of US applicants filing outside the US. They found that non-US applicants conceived longer applications when filing in the US, while US applicants filing outside their country drafted shorter applications. From this the authors concluded that it must be "the American patent system itself that induces a larger size, implying that patentees who apply to the US specify and formulate their claims in much more detail than would be required by the continental European system" (idem). The authors also signal Japan's somewhat special

Figure 3.7. Average size and number of EPO applications according to priority countries



Source: Archontopoulos et al. (2007), based on EPO data. The size of the bubbles reflects the number of EPO applications from the corresponding country.

position – Japanese patents are known to have fewer claims than other countries; however, they have a larger number of pages, indicating that Japanese applicants are asked for longer descriptions of their inventions, which potentially leads to a higher level of information disclosure (idem). All in all, one can conclude that there are large differences between the US patent drafting style and other systems. Van Zeebroeck et al. (2009) found the US style to be a significant and large determinant of patent voluminosity. Furthermore, Archontopoulos et al. (2007) found the so-called "US syndrome" to be even more pronounced with PCT than non-PCT filings. The US drafting style, the PCT route and the effect of common law are the most important factors increasing patent size.

3.2.4. Technological Complexity

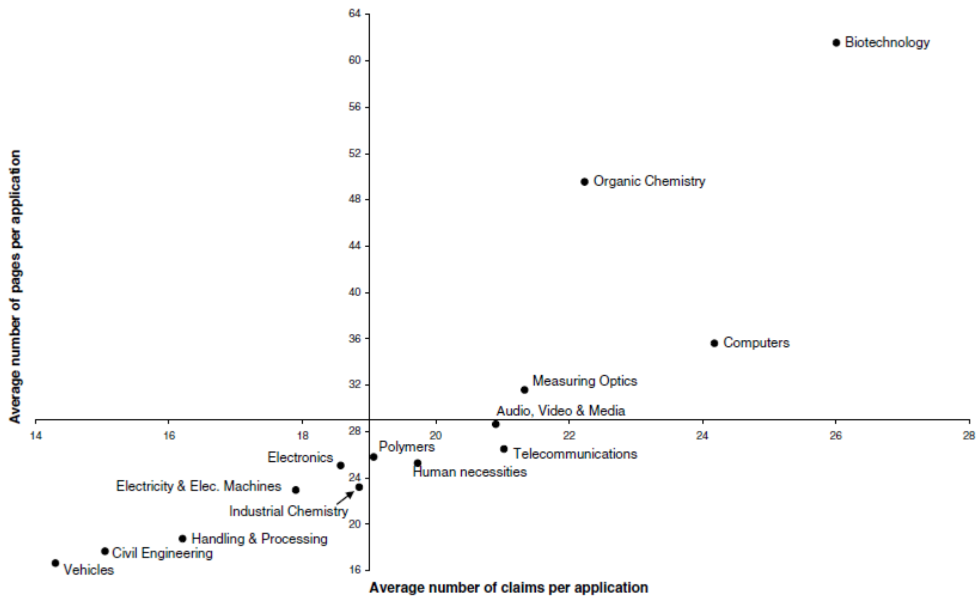
Van Zeebroeck et al. (2009) formulate their hypothesis about the relationship between patent size and technological complexity as follows: "As technology becomes more complex, more words may be required to describe and claim it. Notably because a dwarf standing on a giant's shoulders may really see farther than the giant himself, architectural inventions tend to lead to increasingly complex inventions and technologies over time, produced by larger and larger teams of inventors with complementary skills and expertise. […] It may be expected that the rise in technological complexity will drive the size of subsequent patent applications."

To test this hypothesis, van Zeebroeck et al. (2009) used four variables and examined their effect on patent size: i) the number of inventors specified in the application; ii) the number of International Patent Classification (IPC) classes which correspond to an invention, where a larger number of classes designates a higher technological diversity; iii) backward patent citations, representing "the number of citations made to previous patents, which indicates a larger use of prior patented art"; iv) non-patent citations, "the number of references to the scientific literature included in the document, which is generally used to identify science-based inventions" (idem). All four variables were shown to be "highly significant and positive determinants of patent size" (idem). In the following, I summarize the authors' results with respect to each of the four variables mentioned above. First, the study found that four additional inventors will be associated with one further claim and four extra pages, which means that each inventor potentially brings an important input to the patent. Second, an application which contains three more IPC classes (and is therefore more complex) will also have one more claim and 2.5 more pages. The third result is that building on more patented prior knowledge increases patent size: 5 additional backward patent citations induce one more claim and 0.5 additional pages. Fourth, the same applies to science-based inventions, where 6 supplementary non-patent citations will add one claim and 2 pages (idem). All these results show that technological complexity will indeed increase patent size through its claims and pages.

3.2.5. Emerging Sectors

Technological specificities also play a role for patent drafting practices. The EPO subdivided its applications into 14 general technical areas called "Joint Clusters" (van Zeebroeck et al. 2008). Figure 3.8 presents the size of patent applications as a function of claims and pages for the 14 technological sectors. As in Figure 3.7, one can notice the almost linear relationship between the number of claims and that of pages, indicating that patent size typically grows in both its dimensions depending on the country of origin and on the technical sector. From Figure 3.8 one can derive that the technical areas with the largest patent size are young, emerging sectors: biotechnology, organic chemistry and computers. Van Zeebroeck et al. (2009) provide a few explanations for this phenomenon, which I regard as plausible. They suggest that "the vocabulary of more recent technologies may be less standardized than in more established fields, requiring more detailed descriptions. Emerging technical fields rely more on recent science than older fields, and are based on (as yet) less well-known natural phenomena, which require more explanation than artefacts based on long recognized and accepted mechanisms" (idem). The authors give two examples: biotechnology is based on molecular biology, while software is based on mathematics, operational research and algorithms – most of these fields belong to more recent science. It should be added that the much larger size of biotechnology applications is probably influenced by the incorporation of genetic sequences. Finally, van Zeebroeck et al. (2009)

Figure 3.8. Average size of EPO applications according to technological sector, 2000



Source: van Zeebroeck et al. (2008), based on EPO data.

stress out that "markets where technology is the most important competitive argument" might also contain more licensing and cross-licensing, for which the scope of protection must be defined with greater accuracy, again leading to larger applications (idem). Therefore new, emerging, science-based, highly competitive sectors contribute to the surge in patenting.

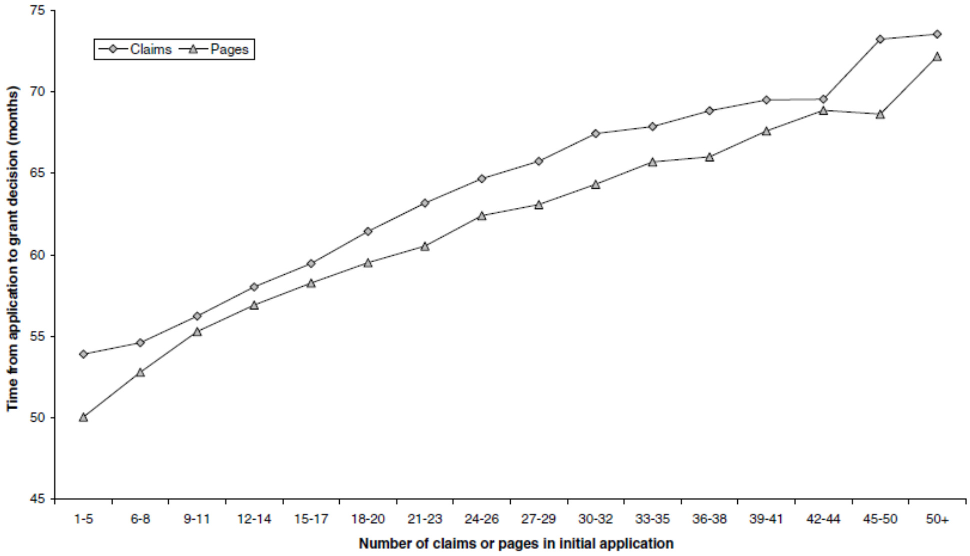
3.3. Effects of the Patent Boom and Possible Solutions

One could see throughout this chapter that patent size has recently experienced a tremendous growth in both pages and claims. This growth, which was not coupled with an increase in R&D activities, can be explained by five factors. The fee structure seemed to have an influence on reducing or compressing the number of claims. Internationalization through the PCT procedure and national practices, which are severely influenced by the US drafting style, were the strongest contributors to patent voluminosity. Following the "draft once, file everywhere" principle, patent applicants typically opt for an American template whenever they want to file an application with international coverage. Belonging to common law, and not to civil law, has a significant effect on patent size. Furthermore, technology plays an essential part: more complex "architectural inventions" depending on more inventors, more IPC classes and more prior art lead to larger applications. The same applies to new, emerging and competitive sectors. Van Zeebroeck et al. (2009) also note that even though these factors explain an important part of the variance in patent size, there remains a significant unexplained trend in the increase of size. They attribute this general trend to "an inexorable path towards more complexity, towards more complete, detailed and hence voluminous literature in every field of activity, such as the user manuals of electronic devices, the documentation of mass-market consumer goods, official or technical reports, or even laws. Patents may be just another playground for this generalized verbosity, encouraged by the decreasing costs of drafting and disseminating written information" (idem).

As stated before, the effects of the patent boom are of utmost social and economic importance. The increasing pressure put on patent offices leads to a questionability of quality standards in the granting process. Excessively large applications "may generate uncertainty in a specific field or indicate that larger areas of knowledge are being appropriated by individual agents" (van Zeebroeck et al., 2008). This uncertainty may compromise the reliability of the patent system altogether, because it means "not only more prior art to sift through for the inventor, patent searcher, patent examiner and the general public, but also a growing scope of the state of the art to invent around or to avoid in order to prevent ultimate litigation" (idem). Stevnsborg and van Pottelsberghe de la Potterie (2007) raise serious concerns about applicants "polluting the technological field" by hiding major inventions in the midst of very large applications. Archontopoulos et al. (2007) also doubt that the number of patent pages is a good indicator for the amount of disclosure of inventions. Overall, it appears that the patent boom might eventually deteriorate the legal certainty of the patent system.

Another important effect is that of seriously delaying the patent granting process. Figure 3.9 presents the average time necessary to reach a decision to grant a patent at the EPO as a function of the number of claims and pages in the application. This figure shows a very strong relation between application size and processing time: the more claims and pages, the larger the delay before granting. This fact "reinforces the idea that larger applications monopolise and consume more resources from patent offices for their processing and hence contribute to the increase in granting delays and backlogs" (Archontopoulos et al., 2007). The problem becomes even bigger when one considers that larger applications are typically filed in technical sectors which already have more applications with existing delays (van Zeebroeck et al., 2008). Figure 3.9 only presents a static view of granting periods. From a dynamic point of view, the situation is even more worrying: Archontopoulos et al. (2007) note that the duration until the granting decision increased over time, from 36 months in the early 1980s to about 57 months in the 2000s. Van Pottelsberghe de la Potterie and François (2006) add an interesting note to this discussion. They suggest that a large chosen number of claims may point out to an intentional choice by patent applicants to postpone the granting process, which comes with high validation fees and translation costs. In my view, a sign of such a strategy is also the increasing share of PCT applications, through which patentees use additional time to raise funds and assess the real value of an invention before incurring the cost of the actual grant. As one final effect, "the growth in document size also results in increasing handling, printing, and shipping costs of patent documents and, when translations are needed, an increase in these costs as well" (van Zeebroeck et al., 2008).

Figure 3.9. Average time to grant decisions at the EPO depending on application size



Source: van Zeebroeck et al. (2008), based on EPO data. Considered were applications filed between 1995 and 1997 and granted by January 31, 2006.

Looking at Figure 3.9, there is one more observation that I would like to make, based on combined information from this figure and earlier parts of the dissertation. In the beginning of Chapter 2.1.3, I mentioned survey evidence by Mansfield (1984), who was reporting that about 60 % of patented innovations were imitated in less than four years after market introduction; moreover, Levin et al. (1987) concluded that almost all patented products are duplicated in less than five years. Figure 3.9 shows us that about 54 months, or 4.5 years, are necessary for an applicant to be granted a patent, if the application is very small and consists of very few claims (1-5). The larger the patent size, the larger is the delay between application time and the granting decision: Figure 3.9 shows that applications comprising of more than 17 claims are granted after more than 60 months, or 5 years, while very large applications with more than 45 claims necessitate more than 72 months, or 6 years, to be granted. In light of the information from Mansfield and Levin et al., one can conclude that for most patents (and especially for medium-sized and large ones), the granting decision takes so long that in the meantime the corresponding patents will have become worthless, since they have been effectively duplicated on the market. This observation raises once more the question about the value and relevance of patents to their applicants.

So can anything be done to reduce or slow down the growth in patent voluminosity? The literature does come up with some potential solutions. Dack and Cohen (2001) proposed a number of measures to be taken, from which I mention the most important: a more rigorous approach to evaluate how clear and precise applications are, the introduction of fees based on the number of claims for PCT applications (which do not exist yet, despite the essential influence of the PCT route on voluminosity) as well as a statutory limitation on the number of claims (in order to circumvent the phenomenon of mega-applications). The EPO has already taken one important step, by introducing a rule (named Rule 29(2) from January 2002) "with the intention to induce applicants to file fewer independent claims within the same category" (Archontopoulos et al., 2007). However, the rule showed only limited impact (van Zeebroeck et al., 2008), and the EPO announced that it needs to review its fees policy to properly respond to the increase in patent size (Pompidou, 2005). For patent offices, the delicate issue in setting claim-based or page-based fees is to "ensure that a point is not crossed when it becomes cheaper to divide an application than pay excess claim fees" (van Zeebroeck et al., 2008).

With respect to the fee structure, I reiterate an idea from Chapter 3.2.1: a large share of applications can be found for 10 claims at the EPO, and for 20 claims at the USPTO. The 10th, the 20th and the 50th claim are the three modes of Figure 3.4. I discuss the first and second mode (10 and 20 claims) in this paragraph, while the next paragraph deals with excess claims (and therefore includes the third mode of 50 claims). The 10th and 20th claims reflect, in my view, a non-linear pricing scheme introduced by patent offices: applicants generate many patent applications with, say, 10 claims, because they would be charged additionally from the 11th claim. Thus, I am raising the question whether a monotone increasing pricing scheme for claims would be better to stop the flood of applications. In other words, instead of charging applicants only for the 10th or 20th claim, patent offices could introduce fees on each additional claim, that is, on the 2nd, the 3rd, the 4th claim, and so on; alternatively, additional fees could be charged in intervals (e.g. for the 3rd, 6th, 9th claim, and so on). However, this might not solve the problem because, in the words of van Zeebroeck et al. (2008), applicants might find ways to break an application into separate parts rather than pay fees for additional claims. My recommendation is to perform further analyses or perhaps empirical investigations to reveal the optimal fee structure.

One approach that proved to work until now was adopted by the USPTO. In December 2004, the USPTO set up a new patent fee schedule which prescribed “significantly higher fees for excess claims and pages”, with the goal of influencing patent drafting practices (van Zeebroeck et al., 2008). The measure was successful and the average number of claims per application dropped from 28 to about 23 claims; moreover, the schedule had an extensive impact on other patent offices worldwide, including the EPO (idem). The main conclusion for this chapter is that the fee structure can be a very useful tool to reduce the explosion of patent applications. However, an optimal fee structure must be designed very intelligently, taking into account all the incentives and trade-offs it might generate.

CHAPTER 4

When Patents Hurt Our Society: Evidence from the Pharmaceutical Industry

"The patenting of drugs and vaccines to treat public health emergencies remains a controversial subject."

Klein (2009)

"Health care professionals who choose the prescription drugs that patients consume may have only attenuated incentives to minimize the cost of drugs to the users."

Caves et al. (1991)

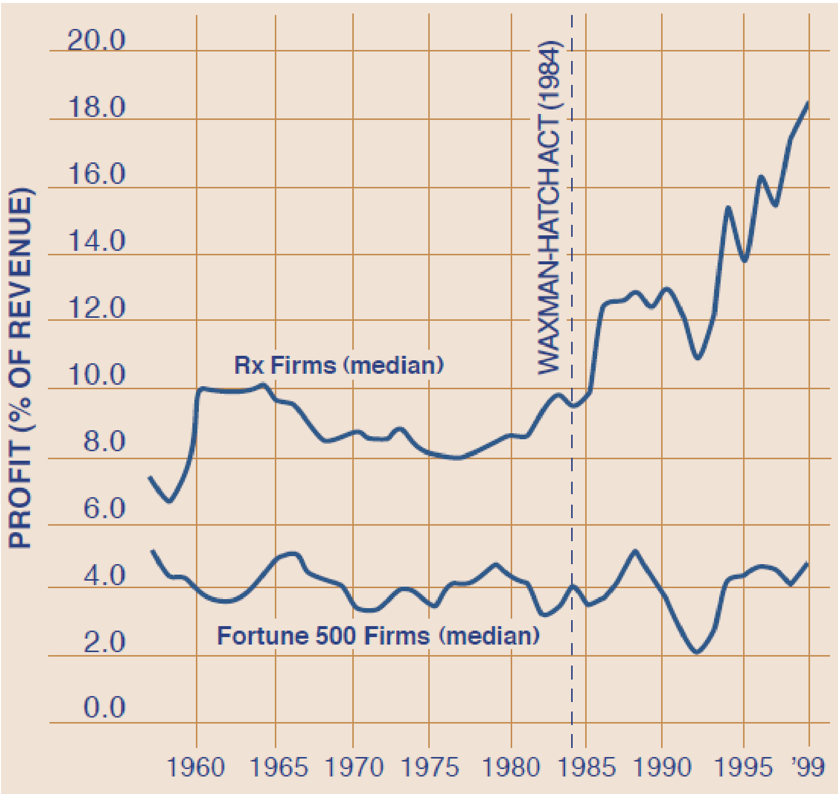
"The mythic costs of R&D are but one part of a larger, dysfunctional system that supports a wealthy, high-tech industry, gives us mostly new medicines with few or no advantages (and serious adverse reactions that have become a leading cause of hospitalization and death), and then persuades doctors that we need these new medicines. It compromises science in the process, and consumes a growing proportion of our money."

Light and Warburton (2011)

4.1. Profits from Pharmaceuticals and the Role of Patents

In the year 2011, the Pharmaceutical Research and Manufacturers of America, or US pharma industry, declared that “Drug patents are good for our health” (PhRMA, 2011). Many researchers and involved actors have challenged this statement and pointed out to different perspectives. The National Institute for Health Care Management Foundation (NIHCM) is a non-for-profit organization who does research on the US health care system. NIHCM (2000) points out that the pharma industry has been the most profitable of all US industries during the last half of the 20th Century, “notwithstanding efforts by both the private and public sector to control health care spending”. Figure 4.1 emphasizes this view.

Figure 4.1. Median profits as percentage of revenues for pharmaceutical companies (Rx Firms) versus Fortune 500 firms



Source: Schondelmeyer (1999), data from Fortune Magazine 1958-2000

One can see in Figure 4.1 that the median profits of pharmaceutical companies (as percentage of revenues) were significantly higher than those of the Fortune 500 companies (the 500 most successful companies in the US) over the period 1960-2000. While the median profits of Fortune 500 firms remained consistently below 6% of revenues, pharmaceutical firms achieved profits between 8-10% of revenues in the period 1960-1984. In 1984, the Hatch-Waxman Act was passed by the US Congress with the goal of encouraging the manufacture of generic drugs (drugs that are chemically identical to brand name drugs, but are marketed under their chemical names and without brand advertising). On the other hand, the Hatch-Waxman Act counterbalanced the promotion of generics by increasing the length of patent protection for brand name drugs, resulting in a surge of profits in the pharmaceutical industry (NIHCM, 2000). Indeed, Figure 4.1 shows that median profits for pharmaceutical companies increased to more than 10% of revenues right after 1984 and continued to increase to levels of more than 18% by the end of the 1990s.

The existence of such overblown profits, combined with concerns about healthcare costs, availability of life-saving drugs and the questionable levels of innovativeness (as described in Chapter 4.2) have led to much criticism of the pharma industry and raised the debate whether the level of pharmaceutical patent protection is at desirable levels.

The US pharma industry's response to such criticism (PhRMA, 2011) was that patent protection is absolutely vital in order to maintain the industry's incentive to innovate and bring any life-saving drugs to the market. The industry suggests that any threat to the existing level of patent protection would be "in no one's interest". The industry's main argument usually revolves around the cost of research and development for drugs: "creating a single new medication costs, on average, about $1 billion", which are funded "from one source alone: private investors" (PhRMA, 2011). The industry maintains that it takes ten to fifteen years of development and testing before a typical drug arrives on the market (PhRMA, 2012).

However, according to different sources, the real figures for the development costs are far, far lower. Angell (2004, p. 46) estimates that the real cost per drug is "well under $100 million". Angell explains how the pharma industry dramatically inflates the cost figures by not considering

R&D tax deductions and by including foregone opportunity costs as out-of-the-pocket costs (opportunity costs represent here the theoretical additional revenue that would have been achieved by investing in a portfolio of financial securities instead of pharmaceutical R&D). Such opportunity costs more than double the cost estimate for drug development (idem). Light and Warburton (2011) point out to other tricks that the pharma industry sometimes uses to increase the cost figures: substantial taxpayer subsidies are not considered, trial costs are inflated, R&D time is exaggerated, and median costs are in fact much lower than the reported mean costs. Besides, the cost reports are not performed by independent institutions, but by centers that are directly funded by the pharma industry and get unique access to the data (Light and Warburton, 2011).

In spite of the profit and cost considerations mentioned above, the conventional view is that patents play a special role in the pharmaceutical industry, especially compared to other industries. This view is endorsed not only by the industry itself (PhRMA, 2011), but also by research studies such as Cohen et al. (2000) and Scherer (2007, 2009). In their study, Cohen et al. (2000) analyzed data from a 1994 survey of R&D managers who were working in 34 industries of the US manufacturing sector. Cohen et al.'s (2000) results showed that R&D managers from the pharmaceutical industry consistently rated patents' importance for the appropriability of profits as higher than did R&D managers of other industries.

One explanation for the alleged higher importance of patents in pharmaceuticals is given by Scherer (2007). He points out to a conundrum faced by the pharmaceutical industry: while the costs of research and development of the chemical formula for a new medicine are astronomical (as one could see above), the costs of actually manufacturing a drug are quite low (once the chemical compound is known, it can be reproduced easily and quickly). So the threat of imitation for pharmaceuticals could be greater and with more severe consequences than it is in other industries. Indeed, the World Health Organization (2012) attests that there exists a large market for copies of drugs that are offered at reduced prices, and counterfeit drugs are proliferated at a very high rate around the world.

Another reason why the pharmaceutical industry is sensitive when it comes to patents can be found in Kesselheim and Mello (2007). They point out that the regulatory approval process that has to respected in order to bring a drug to market requires the disclosure of many aspects of a drug's composition, as well as its development and therapeutic effects. Therefore, the pharmaceutical industry should, at least in theory, rely less on secrecy about the product and it needs to rely more on patent protection. However, there is a contradiction between the conventional view and the industry practice: Cohen et al. (2000) interviewed R&D managers of pharmaceutical companies and found out that they named other instruments as more important for the appropriation of profits. The R&D managers stated that secrecy and first-mover advantage are more important than patents, while patents are used as legal weapons in order to block rivals, to negotiate or to prevent legal suits.

All in all, one finds many contradictions when it comes to the role of patent protection in the pharmaceutical industry. The increase of protection through the Hatch-Waxman Act of 1984 was followed by tremendous increases in profits. The industry is suspected to dramatically inflate development costs and to be secretive about its development data. Even though patent protection is rated as more important in the pharmaceutical industry than in other industries, in reality its role may be overestimated: other instruments seem to guarantee the protection of profits, while patents are often used for legal battles. So it seems to me that the main purpose of patents – that of keeping incentives alive for the industry to innovate – may not be fulfilled in the pharmaceutical industry to the extent that a healthy advanced society wants and needs. The following chapter brings more information which can be useful in the debate whether the rate of patent protection *vis-à-vis* industry innovativeness is optimal or not.

4.2. The Questionable Innovativeness in the Pharmaceutical Industry

In a study from 2002 (NIHCM, 2002a), the National Institute for Health Care Management classified new drugs by their level of innovativeness. The classification for being more or less innovative is based on criteria used by the Food and Drug Administration (FDA). The study by NIHCM (2002a) examines drugs approved for sale in the US market between 1989 and 2000. When a pharmaceutical company applies to the FDA for approval of a new drug, the FDA classifies the drug based on its chemical composition, and on its therapeutic potential. According to NIHCM (2002a), the FDA classifies drugs into three groups:

- new molecular entities (NMEs) are any compounds which have never been approved for sale on the US market. This also includes molecules which differ somewhat from other molecules in the same class of drugs;
- incrementally modified drugs (IMDs) are compounds that already exist on the market, but are newly offered in a different dosage or a different method of administration;
- other drugs are identical in form and composition to existing drugs on the market, but they are newly offered by a different manufacturer.

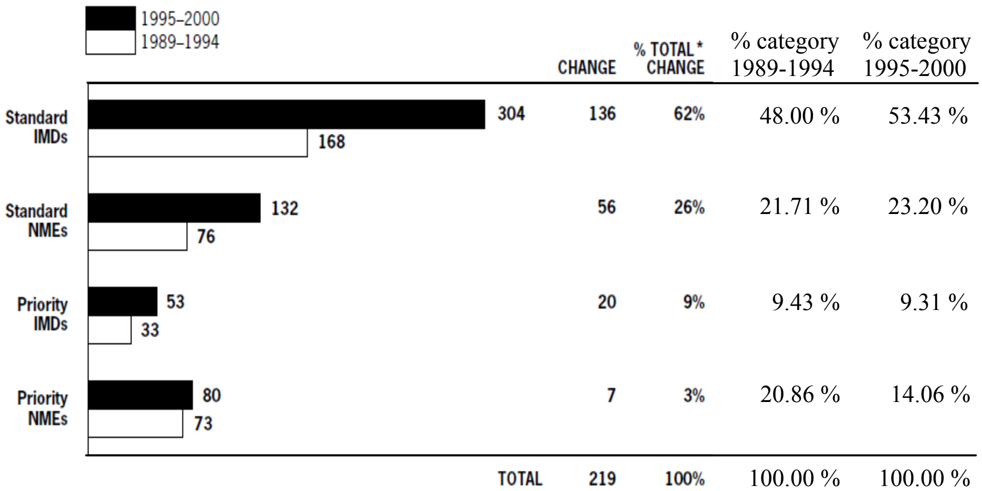
Since I am looking for a measure of innovativeness for the industry, the above classification gives reasons to believe that NMEs are the most innovative (and “the most therapeutically and economically significant” according to DiMasi et al., 1991), IMDs are modifications of existing drugs, while other drugs are just re-packaged drugs (which I can call routine or fake innovations[4]). As stated before, the FDA also classifies drugs according to their therapeutic potential. Drugs showing only a minor therapeutic improvement are considered "standard", while drugs offering a major or a completely new therapeutic improvement are considered "priority" (DiMasi and Faden, 2011). Given the two described dimensions for drug classification (new vs. incremental and standard vs. priority), I obtained four categories of pharmaceutical innovations, with standard IMDs and NMEs as the least innovative drugs, and priority IMDs and NMEs as the most innovative ones.

The data in Figure 4.2 provided by the NIHCM (2002a) depicts the development of innovativeness in the pharmaceutical industry: the least innovative drugs, the standard IMDs and NMEs have had the strongest percentage increase between periods 1989-1994 and 1995-2000, with 62% and 26% respectively. During the same periods, drugs with better therapeutic effects (the priority IMDs) have increased by 9% only, while the most radical innovations (the priority NMEs) increased by a mere 3%.

[4] According to the terminology given by Schade (2013).

However, the computations given by the NIHCM (2002a) have some limitations, because the percentages are based only on the change from the older period (1989-1994) to the newer period (1995-2000) and the base index from the older period was completely neglected by the NIHCM. For these reasons, I have run my own computations of the percentages taken by each of the four categories in each of the periods. These computations are also included in Figure 4.2 in the third and fourth columns.

Figure 4.2. The development of standard and priority NMEs and IMDs between periods 1989-1994 and 1995-2000



*category change divided by total change
Source: NIHCM (2002a). With additional computations by the author in the third and fourth columns

The numbers I have computed myself give a totally different image from that of NIHCM (2002a). Looking at the third and fourth columns of Figure 4.2, one can notice that the development of the four categories has not, by far, been as drastic as in the original interpretation given by the NIHCM. Overall, standard NMEs and priority IMDs have varied very little between the two periods considered: standard NMEs increased very slightly, from a percentage of 21.71 % to 23.20 % of the total, while priority IMDs fell very slightly, from 9.43 % to 9.31 % of the total. On the other hand, there has been a noticeable increase in standard IMDs (from 48 % to 53.43 % of the total), to the detriment of priority NMEs (which fell from 20.86 % to 14.06 % of the total).

Therefore, one could suggest that the innovativeness of the pharmaceutical industry has slightly fallen from the initial period of 1989-1994 to the new period of 1995-2000. However, unlike some critics of the industry would like to argue (see for instance Light and Warburton 2001 in the next section), I found another possible explanation[5] for the apparent decrease in innovativeness (or shift to more incremental innovation), based on my computations from Figure 4.2. Namely, it could be that there are some “true innovators” who produce mostly priority NMEs, but there are much more “copycats” who only produce standard IMDs. Based on the data I have in this dissertation, it is hard to disentangle the true explanation for the questionable innovativeness in pharmaceuticals. However, the following Chapters 4.3 and 4.5 also point out to some intuition that it has become increasingly difficult for “true innovators” to come up with radical medicine, and that the market for the proliferation of “imitations” has been flourishing lately. Thus, it is hard to tell whether patent protection should be increased or decreased. If it is the case that imitation has become the more prominent type of “innovation”, patent protection for radical pharmaceutical innovation should be enhanced, and not reduced.

In this section of the current subchapter, I provide some of the literature which is critical of the pharmaceutical industry. Light and Warburton (2001) write a comment which criticizes the pharma industry and accuses it of falling levels of innovativeness: “R&D costs need not be such an insuperable obstacle to the development of better medicines. The deeper problem is that current incentives reward companies to develop mainly new medicines of little advantage and compete for market share at high prices, rather than to develop clinically superior medicines with public funding so that prices could be much lower and risks to companies lower as well”.

Contributing to the criticism, the NIHCM (2002a) study directly attributes the rise in IMDs to the strengthening of intellectual property protection, through the Hatch-Waxman Act from 1984. The act allows for patents to be extended by another three years, if the FDA approves a "new use" for an existing patented drug (idem). Furthermore, the act extended the length of

[5] I am thankful to Christian D. Schade for suggesting this interpretation.

patent protection for brand name drugs, giving them up to five more years of protection (NIHCM, 2000, pp. 4-5; Scherer, 2009, p. 197). Other sources (e.g. The New York Times, 2011) express their concern that the major pharmaceutical companies have very few promising new drugs under development, and try to extend protection of older drugs for as long as possible. On the other end of the spectrum, there are researchers such as Grabowski and Wang (2006), who insist that the pharmaceutical industry remains very innovative. They point out that simple counts of NMEs are misleading and that further analyses would give a different picture, according to which very high-quality radical medicine continues to be introduced on a large scale.

Overall, it is very difficult to understand whether the industry's innovativeness has fallen or has remained at desirable levels. The literature remains controversial on this topic, with many arguments left on both sides. Should an entity such as a patent regulator understand which level of protection is optimal to encourage innovativeness, questions still remain about the most effective regulating instrument for pharmaceuticals. As we have seen in Chapter 2, regulators could rely on instruments such as patent length, height or breadth. I believe that the optimal choice of a regulatory instrument would be a very hard decision. Patent length is the easiest to implement, for it is quite straight-forward: it would just imply setting a shorter or longer lifetime of a patent. Patent height, defined in Chapter 2 as stronger novelty requirements, or as restrictions on the improvement of patented products, could work for pharmaceuticals, especially for the IMDs: here, the regulator could reduce the spread of IMDs by setting harsher (higher) conditions for patenting. Patent breadth would be the most difficult regulating instrument. One can recall Table 2.1, which synthesizes the myriad of definitions for this instrument as well as Lerner (2009, p. 343) who recognized how difficult it is to measure and interpret patent breadth. Based on the definitions from Table 2.1, I believe that patent breadth is the most delicate instrument: for example, setting a large breadth for a new molecule (the position of a chemical radical or structure of an isotope) might block competitors from innovating around the patented molecule, and thus it may reduce the future innovativeness of a pharmaceutical compound or class of compounds.

4.3. Competition with Follow-on Drugs

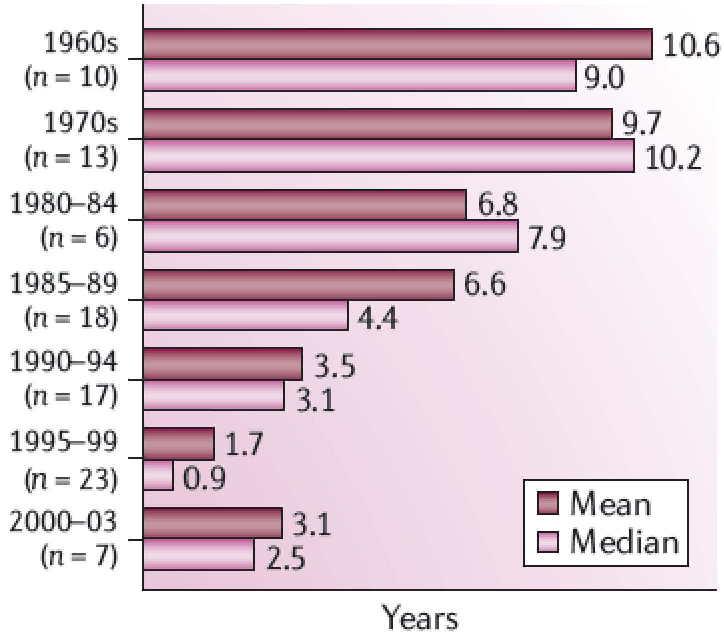
So, most of the innovative activity in the pharmaceutical industry is based on incremental innovations, also called “follow-on” drugs or “me too” drugs in the literature (e.g. in DiMasi and Faden, 2011). This chapter describes the nature of competition for “me too” drugs.

A “follow-on” or “me too” drug is generally defined as “a drug with a similar chemical structure or the same mechanism of action as a drug that is already marketed” (idem). Such slightly modified drugs have been the subject of much controversy. Some argue that follow-on drugs “often provide useful alternatives or enhanced therapeutic options for particular patients or patient subpopulations, as well as introduce price competition” (idem). Critics of “me too” drugs consider that only drugs which make it to the market first (“first-in-class” drugs) represent genuine and valuable innovation, while “me too” drugs lead to duplication of R&D costs and to a waste of resources that could be better directed elsewhere (Angell, 2004, Hollis, 2005).

DiMasi and Faden (2011) analyze the timing of patent filing and the entry of products in 94 therapeutic classes on the United States market between 1960 and 2007. In each therapeutic class, they took into consideration the introduction of the first-in-class compound, as well as that of follow-on drugs (on average, in each class, about 4 follow-on drugs were introduced after the first-in-class). In the following, I will look at their results on developments of the speed and quality of innovation over time.

Figure 4.3. shows that, over the last four decades of the 20th century, the mean length of the period of marketing exclusivity for first entrants in a US therapeutic class declined substantially, falling by about 84% from 10.6 to 1.7 years from the 1960s to the end of the 1990s. A decreased marketing exclusivity is an indicator that it became easier for imitators to bring follow-on pharmaceutical products to the market. DiMasi and Faden (2011) run regressions showing that the speed of market entry for competitors to first-in-class drugs increased by about 3 years per decade throughout the period from 1970 to 1999. This fact contributes to the discussion from Chapter 4.2, where we have seen a large growth of IMDs. Here one can notice consistent reasons for the growth on incremental innovation, to the detriment of radical

Figure 4.3. Period of marketing exclusivity for first entrants to a therapeutic class in the US (time from first-in-class approval to first follow-on approval)



Source: DiMasi and Faden (2011)

innovation (in the form of NMEs): the decreasing market exclusivity made it more likely for "copycats" to access the market, and more difficult for "true innovators" to hold on to their market status as radical innovators. For such reasons, patent protection for "true innovators" could and should be enhanced.

So competition in pharmaceuticals did increase (in the form of the speed of market entry), but was it all competition of purely incremental innovations offering only minor improvements over existing treatments? A previous study by DiMasi and Pacquette (2004), confirmed by DiMasi and Faden (2011) provide a surprising answer to this question. Only about one-third of all follow-on drugs (32% precisely) received a priority rating by the FDA. Such a priority rating is only given to drugs delivering significant advances in treatment or offering treatment to a condition for which no adequate therapy exists. Consequently, DiMasi and Faden conclude that the first-in-class drug is not necessary the best, and there is innovation of considerable value in follow-on drugs. However, DiMasi and Faden's argument can be reversed. The fact that 32 % of follow-on drugs received a priority status implies that the remaining 68 % *did not* receive the priority rating. From a simplified perspective, one could regard the matter as follows: only about 32 % of pharmaceutical manufacturers seem to be somewhat alike "true innovators", while the other 68 % seem to resemble "copycats", which strengthens the argument for increased patent protection.

Competition within a therapeutic class also has some effect on drug prices, although this effect is not entirely intuitive. Lu and Comanor (1998) show that drugs with advanced therapeutic effects (class A and B drugs) are sold for "substantial price premiums over their existing substitutes", and these high prices only drop slowly over time. Drugs offering little improvements over existing substitutes (class C drugs) enter the market at lower prices but, paradoxically, these prices increase substantially during the time after launch. Still, both entry prices and subsequent increases "are lower when there are more competing products" on the market (Lu and Comanor, 1998).

We have seen in this section that the speed of market entry for follow-on drugs substantially diminished during the last decades of the 20th century, there was increased competitiveness in all drug classes, and many follow-on drugs represented innovations of considerable quality. Moreover, competition slightly affected drug prices. DiMasi and Faden (2011) provide several reasons for their results, some of which are: "increased opportunities from advances in biomedical science; shifts in drug development approaches that make connectedness to scientific networks more important; legislation making generic entry easier; […] expansion of prescription drug insurance to a larger segment of the population".

4.4. Patent Races in Pharmaceuticals

In this chapter I first look at racing behavior among pharmaceutical companies for the attainment of patents, and briefly contrast it with considerations from theoretical models. Second, I present empirical evidence for patent races in pharmaceuticals.

When it comes to patent races, one of the main concerns in the theoretical literature is that the racing firms may be involved in wasteful duplication of R&D efforts, and many resources could have been allocated to more productive purposes (Scotchmer, 2004). But is this really the case in practice, and do firms indeed compete to the extent that rent dissipation occurs in the pharmaceutical industry?

A study by Cockburn and Henderson (1994) suggests that such a wasteful behavior does not occur. They analyze the modern game theoretic literature on patent races (which will be extensively described in Chapter 5.1) and try to match empirical findings to the predictions of this literature. They first motivate the selection of the pharmaceutical industry for their study: the industry "has often been held up as a prime example of the types of strategic racing behavior predicted by the literature" (idem). The reason for the relevance of pharmaceuticals is the industry's high degree of technological competition with heavy investments in R&D (idem). These investments are necessary since "successful research is a key contributor to commercial success" (idem), yet they are subject to extremely high uncertainty: DiMasi et al. (1991) estimate the success rate of clinically approved chemical compounds to be around 23%, but they point out that a very small portion of investigated compounds enter clinical trials at all.

Even though the pharmaceutical industry is highly relevant in the search for patent races, Cockburn and Henderson (1994) criticize the low applicability of the game theoretic patent race literature for the typical “multi-product multi-project” R&D firm, due to some critical assumptions: the so-called "winner-take-all" setting present in most patent race models, and the deterministic R&D function fundamental to many of these models. Cockburn and Henderson also state that testing the predictions of this literature cannot be done without the use of "heroic assumptions" (e.g. total appropriability of profits and consumer surplus, lack of knowledge spillovers or perfect substitution of competing projects). They also show that relaxing these assumptions usually leads to indeterminate results: “competitive industries may invest too much, too little or just about the right amount in research” (idem).

Cockburn and Henderson (1994) “use unusually detailed data on research investments and outcomes gathered at the level of individual research programs conducted within ten pharmaceutical firms over a period of more than seventeen years”. The chosen firms include the most important American and European manufacturers, account for 25-30% of worldwide

R&D and sales, and add up to about 2.700 observations indexed by firm, research program and year.

Cockburn and Henderson run simple correlations in R&D investments and outputs across firms, focusing on drug discovery and not on drug development, since the first is much more relevant for R&D activity. The authors find very weak correlations between investment levels across firms (which become insignificant when including measures of technological opportunity and demand shocks), but very high and significant autocorrelations of investment levels within the same firm. They interpret these results as proof of significant unobserved firm heterogeneity in investment patterns and they reject the hypothesis that firms are entrenched in cut-throat patent races or "tit-for-tat" strategies. According to the study, competition for a single prize (the "winner-take-all" setting central to the patent race literature) appears to be an inappropriate description for patent races in pharmaceuticals, as does the "me too" (or responsive) investment. Instead, the authors find that competing R&D projects across firms seem to be complementary, while "the industry is characterized by substantial spillovers of knowledge and similar research can lead to related but significantly different outcomes" (idem). Moreover, drivers of investment decisions are better characterized by "the heterogeneous capabilities of the firm, by adjustment costs, and by the evolution of technological opportunity" (idem).

It might be true that the patent race literature has some weaknesses, in that it over-simplifies reality. It might also hold that the literature should accommodate elements typical for multi-project multi-product R&D firms, such as multiple prizes in R&D competition. However, a patent race model which would incorporate so many details would lead to an extreme surge in mathematical complexity and could not be delivered to subjects experimentally. In order to develop any patent race model of manageable complexity (and, even more difficult, a model feasible for experimental testing), one cannot easily expect to generate models of multi-project multi-product heterogeneous R&D firms with multiple prizes. Instead, I believe that one has to keep the critical assumptions of "winner-take-all" and deterministic (or stochastic) R&D functions, and focus on a one-project, one-product heterogeneous R&D firms with a single prize. This assumption is not necessarily too hard. One can imagine isolating a single product or project that firms compete on, and look away from all the complexities of a multi-project competition, with all the interactions it involves. Indeed, in my experimental Chapter 6 I keep these vital assumptions from the patent race literature, in order to prevent an already given high complexity from reaching extreme levels. Perhaps it is so that Cockburn and Henderson's (1994) expectations are too much to ask, and that their multi-project, multi-prize assumptions simply do not apply in the context of a more or less "simple" patent race model. Or, maybe these authors' findings are only valid for the pharmaceutical industry and different results would be obtained in other industries.

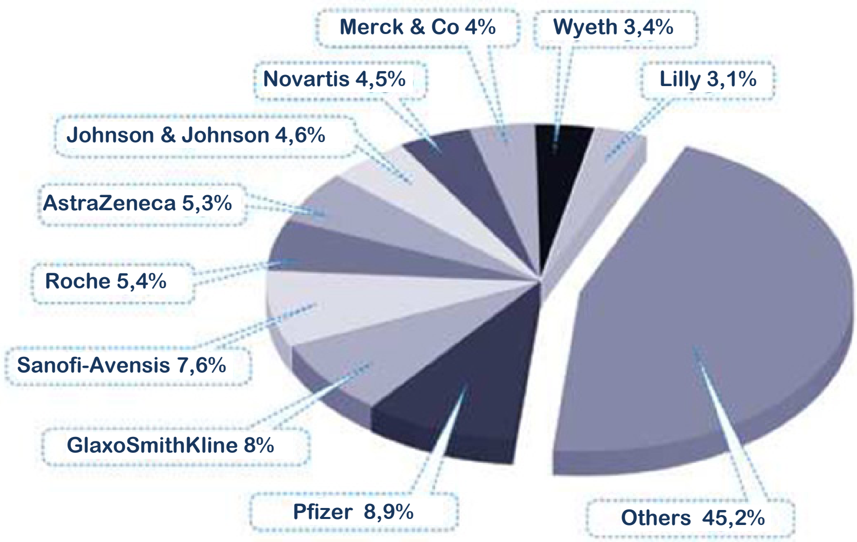
DiMasi and Faden (2011) also investigate the nature of racing behavior in pharmaceuticals. They ask the question whether firms engage in patent races or mostly in imitation. In order to answer this question, DiMasi and Faden analyze data from the Tufts Center for the Study of Drug Development (which maintains databases of new drugs and biopharmaceuticals approved in the US). The authors find that "all of the follow-on drugs for classes in which the first-in-class drug was approved from the late 1980s onwards were synthesized before the first-in-class drug was approved. Indeed, for classes in which the first member was approved since 1990, 80% or more of the follow-on drugs were in clinical trials somewhere in the world before the approval of the first-in-class drug" (idem). Even more, "for drug classes in which the first-in-class drug had been approved since the 1970s, 90% had at least one follow-on drug in its class with its first worldwide patent filed before the first-in-class drug was approved" (idem). DiMasi and Faden conclude that these figures are proof for a race to market among drugs in a therapeutic class, and not for low-risk imitations of already marketed breakthrough drugs.

Combining the results by Cockburn and Henderson (1994) and those of DiMasi and Faden (2011), I can conclude that racing behavior with pharmaceuticals exists to some degree, but that patent races are not of the "winner-take-all" type which concerned researchers due to the duplication of R&D efforts. Therefore, this assumption simply does not seem to hold in the context of pharmaceuticals. Instead, in such a context the market accommodates several prizes for several market players. Most importantly, Cockburn and Henderson interviewed several R&D managers from the pharmaceutical industry about their research strategy. The managers answered that they mostly avoided head to head competition, and that they generally viewed racing behavior as highly unproductive. Instead, the managers employed other criteria when making their decisions: "the size of the unmet medical need, the scientific potential of a field, and the idiosyncratic capabilities of their researchers", as well as the "views of their marketing departments" (idem). So, firms try to avoid strong racing and settle to the sharing of the market, resulting not in one single winner, but in several winners.

Figure 4.4 gives information that is relevant to the subsequent discussion about racing behavior with pharmaceuticals. The figure presents the ten most important players in the pharmaceutical industry. One can see that Pfizer and GlaxoSmithKline are the worldwide market leaders, and that about 55% of the world revenue for pharmaceuticals is concentrated in the hands of the top ten companies.

Several empirical examples are in tune with the conclusion formulated above – that races result in several winners, not just in one winner. A first example is the racing behavior for the development of statins, which are drugs that lower cholesterol levels. Table 4.1 shows that, over time, there

Figure 4.4. Market shares for the ten largest pharmaceutical companies worldwide



Source: ETC Group (2008)

were races to market between the main industry actors. However, in the end the market settled to accommodate more than one player (or "winner").

In Table 4.1, I found significant competition and racing behavior in the case of statins. Merck cannibalized its own market by introducing Zocor, a more potent statin than its previous product, Mevacor. Pfizer's drug, Lipitor, which came to market after the approval and marketing of four previous drugs in the same therapeutic class, became the best-selling drug in history (Simons, 2003), showing again that follow-on drugs may provide valuable innovation. At the other end of the quality spectrum, Bayer's 1998 drug Baycol was later withdrawn after increasing evidence that it causes muscle failure, resulting in at least 52 deaths (Cable News Network, 2001). So, there are significant differences between drugs in the same class, as illustrated by the example of statins, which were also shown to have differing effects for different patient populations in a study by Chong et al. (2001). The "winner-take-all" setting could not be confirmed in the case of statins: if one winner "took it all", only one statin would have survived on the market; instead, statins continue to co-exist up to this day.

A second example for the mildness of "racing", re-enforcing the conclusion I stated above that the market allows for several "winners", is the case of blood pressure lowering drugs, whose development I present in Table 4.2. At first glance, it might seem from Table 4.2 that firms engage in patent races. The fact that Squibb and Merck with their drugs Captopril and Enalapril held 82% of the US market in 1990 (Cockburn and Henderson,

Table 4.1. Racing behavior for the development of statins

Introduction Year	Producing Firm	Drug Name
1987	Merck	Mevacor
1991	Sankyo	Pravachol
1991	Merck	Zocor
1994	Novartis	Lescol
1997	Pfizer	Lipitor
1998	Bayer	Baycol

Source: Simons (2003)

1994) indicates that the other nine compounds from Table 4.2 could be seen as evidence for wasteful over-investment by the losers of the race. But a more profound exploration of the further developments of these drugs shows a different reality. In fact, many of the companies which were initially involved in developing blood pressure lowering drugs continued to engage in “fundamental research and open exchange of information” (Patchett et al., 1980). They continued to publish actively, even at rates exceeding those for the patented drugs listed in Table 4.2 (Cockburn and Henderson, 1994). Some of these companies formed joint ventures and many developed alternative therapies (idem).

My general conclusion for this subchapter is that there are strong patent races and competition among pharmaceutical companies, but no "winner-take-all" setting and no wasteful duplication of efforts. Instead, firms avoid cut-throat competition, they make use of their initial investments, develop further projects based on their gained knowledge, and may even co-operate. Knowledge spillovers flow throughout the industry, and the market has enough room to accommodate several winners.

Table 4.2. Racing behavior for the development of blood pressure lowering drugs

Patenting Year	**Producing Firm**	**Drug Name**
1977	Squibb	Captopril
1980	Merck	Enalapril
1980	Merck	Lisinopril
1980	Dainippon	Alacepril
1982	French public sector	Perindopril
1982	Warner Lambert	Quinapril
1982	Squibb	Fosinopril
1982	Chugai	Moveltipril
1982	Takeda	Delapril
1983	Hoechst	Ramipril
1983	Hoffman La Roche	Cilazapril

Source: Cockburn and Henderson (1994)

4.5. Generics: Beyond the Death of Patents

The purpose of generic drugs is to replicate the therapeutic effects of brand name drugs, but at lower prices, thereby increasing people's access to medicine. Producers of generics in the US must respect strict requirements by the FDA in order to get approved – they must show that their generic drugs are bioequivalent to the brand name drugs (NIHCM, 2002b). Bioequivalence means that the generic must contain identical active ingredients, in the same amount, and with the same effects on the body as the innovator drug (idem). NIHCM (2002b, p.10) writes that "today, virtually all generic drugs are essentially the same as the innovator drugs they copy". A generic drug is only allowed to enter the market after the patent for the brand name drug has expired. Since generics' producers do not have to duplicate the R&D effort – the most significant cost factor in a drug – but instead they only have to incur manufacturing costs, generics can be offered at greatly reduced prices. It is therefore expected that generic drugs entering price competition should quickly conquer the market and drive down prices after the original patent expires. However, the current chapter will show that this is not always the case, and that generic competition contains some unexpected complexities.

As stated before in Chapter 4.1, the Hatch-Waxman Act of 1984 was passed with the purpose of reducing barriers to generic competition. NIHCM (2000) explains the effects of the Hatch-Waxman Act. Before the Act, manufacturers of generics had to undergo "expensive clinical trials to prove the product's safety and effectiveness" (idem). Moreover, they could not perform any generic development while the original drug was still under patent protection. After Hatch-Waxman, generic companies were allowed to forego full clinical trials if they could prove that the generic was bioequivalent to the brand name drug. In addition, the Act permitted generics' manufacturers "to make or use a patented product, perform all necessary testing, submit an application and even receive tentative approval before the relevant patents on the originator drug expire" (idem). Thus, a manufacturer could "bring its product to market on the very day that the branded drug loses its protection" (idem), dramatically decreasing the time between patent expiration and generic market entry. On the other hand, to compensate for this decrease, brand name drugs received extended patent protection for several years (as pointed out in Chapter 4.1).

The study by Magazzini et al. (2004) examines generic competition in four countries: the US, the UK, Germany and France. The study analyzes the development of generics' market shares and prices in the four countries by using sales data for drugs with expired patents between 1986 and 1996. Magazzini et al. found that market penetration rates for generics were highest in the US and lower in the European countries. They attributed this phenomenon to the amount of price regulation in different countries – strong price regulation (as in Europe) discourages generic entry, while weak regulation (as in the US) promotes generics by allowing higher profits to be made. However, even though generic penetration is higher in the US, drug prices follow a very surprising development. Magazzini et al. show that, in Germany, the UK and France, prices of brand name drugs steadily decrease after generic entry and eventually converge with generic prices. Entirely against expectations, in the US, prices for brand name drugs *increase* after patent expiration!

So what explains the paradox of rising prices for brand name drugs after patent expiration in the US? Magazzini et al. (2004) provide a part of the answer: many European countries have nationalized or heavily regulated health care systems and have implemented many regulatory measures designed to keep the cost of health care under control. In Germany, insurance companies have to take into account a reference price system when they make drug purchases, and there is a ceiling on total pharmaceutical expenditures (Huttin, 2002, p. 86). In the UK, doctors are given a fixed budget for drug expenditures and this makes them very conscious about drug prices (idem, p. 82). France also has a national budget for drugs and, above all, it regulates drug prices directly (idem, p. 81).

The study by Caves et al. (1991) gives the other part of the answer to the question above. Caves et al. examined prices, market shares and advertising for thirty US drugs that lost patent protection between 1976 and 1987. They showed that an essentially different mindset reigns in the US compared to Europe. In the US, the health care professionals who choose which drugs should be prescribed for their patients do not have the incentives to minimize the cost of drugs to the users, or to their insurers. Instead, there exists a misalignment between the interests of patients and the incentives of rent-seeking pharmaceutical companies. While consumers are sensitive to prices (because more expensive drugs are reflected in higher health insurance premiums), US physicians often have little or no information about drug prices (in stark contrast with European systems). Physicians' behavior is largely influenced by extensive sales promotions by pharmaceutical firms, which also include direct visits to health care professionals designed to inform them about new drugs and increase their brand loyalty. Caves et al.'s study found that generic drugs in the US were sold at an average of about 60% of the price for the original brand name drugs from 1976-1987. Since this represents an important discount, one would expect brand name drug prices to fall and converge to generic levels, just like in Europe. However, this is not what the study found – brand name market share lost to generic competition was quite low and generic penetration was only 36% in 1987. Moreover, new generic entrants do not seem to affect prices for brand name drugs but instead they only have an influence on the prices of already existing generics. The authors' conclusion is that the intensive marketing practiced by pharmaceutical firms generates brand allegiance and goodwill in medical professionals, who, in the absence of information about prices and influenced by a "business as usual" mentality, prescribe the more expensive drugs to their patients.

These factors explain how prices for brand name drugs could increase in spite of the market entry by generic competitors. Marketing efforts by the pharma industry maintain a strong price differentiation between brand name and generic drugs even after patent protection has expired. This has two important implications. Firstly, it means that the initial goal of generic competition which I have stated at the beginning of the subchapter – offering the same therapeutic effects at lower prices – might not really be fulfilled in the US, or not to the extent that would benefit most people. Secondly, it means that pharmaceutical firms have well-functioning alternatives to patent protection: after the death of patents, marketing measures are effectively used to maintain high market shares for brand name drugs.

4.6. Conclusions and Recommendations for Patents and Pharmaceuticals

As Chapter 4 shows, the market for pharmaceuticals is fairly complex and contains many unexpected intricacies. Generally, it is unclear whether patent protection is at optimal levels, and the effect of patents on industry innovativeness is the subject of hot debates in the literature. It is furthermore questionable whether patents benefit the industry or society.

The pharma industry was the most profitable of all US industries during the second half of the 20th Century, with median profits significantly higher than the average Fortune 500 companies. Attempts to encourage generic competition such as the 1984 Hatch-Waxman Act were somewhat unsuccessful: market penetration for generics remained low, brand name market shares stayed high and prices for brand name drugs even increased due to marketing efforts by pharmaceutical firms. The Hatch-Waxman Act also had unintended consequences for society, as the amount of additional patent protection for pharmaceuticals seems to have led to an explosion of industry profits. In spite of the existence of such large profits, it remains questionable whether the innovativeness of the industry has increased or not. Based on the limited data I have gathered in this dissertation, it seems that there has been a shift from radical innovation (in the form of new molecular entities) to more incremental medicine. However, this might be the result of an increased easiness of "copycats" to access the market, at the detriment of "true innovators" of medicine. The most competition seems to take place with incremental innovations, or "me too" drugs. There are signs of patent races in the industry, but racing behavior is not of the "winner-take-all" type described in the patent race literature; instead, the market accommodates several winners who have no interest in cut-throat competition.

Last but not least, patents' role for pharmaceuticals seems to be somewhat overrated. R&D managers named different protection instruments (secrecy and first-mover advantage) as more important than patents, while patents are used as legal weapons. Furthermore, even after the death of patents, the pharma industry still performs very well by building brand loyalty and strong bonds with medical professionals, which again shows that patents might not be of essential importance to protect market share.

My recommendations for the regulation of the pharma industry are as follows. Patent regulators should decide whether or not "true innovators" should be protected agains "copycats" and consequently the levels of patent protection can be enhanced or reduced. There should be increased taxation on the industry's profits. In the US, third-party payers such as state reimbursement programs and private insurance should put pressure on medical professionals (or at least inform them better), so that more generic medicine can be prescribed. Ethical considerations should be brought into the foreground – medicine is not a regular product; instead, drugs save lives and public policy should do its best to guarantee access to drugs for as many people as possible. Finally, the industry's innovativeness should be spurred by concentrating on smaller firms: Fershtman and Markovich (2010) write that "in the pharmaceutical market small startups are typically more efficient in the development of a new drug than the big pharmaceutical companies". Therefore, a shift towards small entrepreneurial firms might help to encourage the pharma industry's innovativeness.

CHAPTER 5

Patent Races and Their Problems

“Roughly speaking, early introduction time can be made still earlier by raising the benefits of being first, reducing the benefits going to imitators, increasing the rivals' innovation rate *h* or reducing the rivals' imitation rate *k*. The preceding statement needs to be qualified, as changes in many of the parameters affect the firm's behavior in ambiguous ways. This ambiguity is consistent with the findings of Scherer [1967a, A/N]; the consequence of a parametric change upon the choice of introduction time will often depend on all the circumstances.”

Kamien and Schwartz (1972)

“Conclusions from these types of models are also rather fragile: small changes in the timing of moves, the information structure of the game or the treatment of spillovers can easily reverse or weaken any given theoretical result.”

Cockburn and Henderson (1994)

5.1. Economic Models of Patent Races. When Models Become Too Complex

The long, complicated and, as we will see, largely unresolved odyssey of the patent race literature starts in the late 1960s, with a prodigious creation of "traditional" theoretical models throughout the 1970s and especially '80s, a sign of research fatigue in the '90s[6], and the development of model extensions and further discussion after the year 2000.

One issue does need clarification before I can proceed with the analysis of the patent literature: where is the boundary between models of optimal patent regulation (which I have extensively examined in Chapter 2), and models of (optimal) patent races? One hint to help us draw a line is the review by Reinganum (1989), which has been catalogued as "excellent" in virtually any subsequent study on patent races. Most of the patent race papers quoted by Reinganum (1989) are analyzed in this chapter. However, in my view, the scientific delimitation between the two types of models comes from the fact that models of patent regulation deal with issues that can be influenced by policy-makers: patent height, breadth and length, as well as their macroeconomic and social effects. By contrast, patent race models are applied to isolated industries with *n* number of firms, in which agents compete for the earliest possible introduction date of their innovations and rewards are given by success in the market. As Judd et al. (2012) put it, models of patent regulation endogenize patent policy but abstract from the R&D competition (the race) itself. Models of patent races take patent policy as given (or exogenous) and focus on the competition in itself.

5.1.1. The "Traditional" Patent Race Literature

Yoram Barzel (1968) opens the floor with his question about the optimal timing of innovations in a competitive industry. It is this question which sparkled later research and eventually led to the development of the patent race literature. In creating his model, Barzel (1968) builds on Kenneth Arrow's (1962) concept of "learning by doing". Both these papers were

[6] Most probably after Reinganum's (1989) groundbreaking literature review.

pioneering at the time: previously, the literature had treated technical change as exogenous to the economic system and imposed on the production function of firms. So Barzel's contribution is twofold: it endogenizes the production of innovations (as did Arrow), and it opens the way for patent races. According to Barzel, innovators' behavior in an industry is under the pressure of two opposing forces. On one hand, competition within the industry leads to firms' necessity to be the first to innovate and generate premature discoveries or introductions of new products. To accomplish this, large resources need to be devoted to innovating activity. On the other hand, firms are often unable to reap the full benefits from their innovations, which leads to under-investment. Therefore, the date of innovation introduction will diverge from the social optimum, depending on several market conditions which are elucidated below. The most important working assumptions of Barzel's model are: a) constant capital cost of developing an innovation; b) growing demand; c) reduction of unit costs at constant quality through the innovation; d) a royalty rate per unit of output for the innovator. After performing a profit maximization problem for the firm, there are two possible outcomes in the industry: either there is monopoly and the innovator captures all the benefits from the innovation (in which case there is no economic growth), or there is competition and the benefits are diffused among several market players. In the latter case the innovation is introduced too early or too late, but there is a catch: growth is only positive if the innovation is introduced too late. In the case of early introduction, the rate of growth falls back to zero (the value of the supramarginal project, the profit, is reduced to zero, due to the higher cost of faster development). I conclude from Barzel's model that society (or the industry) can only benefit from the cost-reducing effects of an innovation if it accepts a lower speed of the innovation.

Kamien and Schwartz (1972) improve on Barzel's (1968) model. They also discuss the choice of development period and introduction time of innovations by firms under competition, compared to a situation without competition. Their assumptions - or factors taken into account by the profit maximizing firm - are: a) increasing cost with the compression of the development period; b) reduction of profit opportunities with the prolongation of development period; c) a subjective probability of rival innovation or imitation. Assumption c) represents the main departure of Kamien and Schwartz (1972) from Barzel's paper. Barzel treats rivalry according to game theoretic and reaction curve considerations on oligopolistic markets with perfect information (each firm knows what the other players want to do). Kamien and Schwartz (1972) add stochasticity to the model via “a single subjective probability distribution over the introduction date of any rival product” (idem). This element of stochasticity is in tune with the uncertainty-related nature of most research and development activity (Nelson et al. 1967), and therefore we can consider Kamien and Schwartz's (1972) paper an improvement over Barzel's. Kamien and Schwartz (1972) set up the firm's choice problem mathematically and show their solution properties. They state that their results contradict Barzel's findings: under intense competition firms will not develop too early (they will not drive supramarginal values to zero), but they will rather postpone development indefinitely, or even drop out completely; only when the conditional probability of rival introduction is low enough, “driven below the rate of growth of the benefits stream will an expected profits maximizing firm deem it worthwhile to embark upon development” (Kamien and Schwartz, 1972). Generally, a firm's behavior will be under the pressure of two forces: on one hand, a firm will want to reduce its R&D costs by pursuing a longer lived development strategy; on the other hand, waiting too long will increase the probability that a rival will be first to introduce the innovation. The second effect, according to Kamien and Schwartz (1972), opens up the possibility that there might be parameter settings in which early introduction is possible, depending on the possibility of imitation by rivals. They acknowledge that "changes in many of the parameters affect the firm's behavior in ambiguous ways" (idem). Kamien and Schwartz also find this ambiguity to be consistent with the findings by Scherer (1967a), and thus the former authors suggest that in models of competition with R&D "the consequence of a parametric change upon the choice of introduction time will often depend on all the circumstances".

So if I were to draw an early conclusion after the analysis of only two patent race models, it would be one similar to the conclusion I will actually draw later on (when models get incredibly hairy): even the simplest patent race models are very sensitive to changes in assumptions and parameters, and results can change drastically by only slightly tweaking on any of the model assumptions. This conclusion also leads to the question: how would real people react when faced with conditions resembling those assumptions?

And this question gave motivation to perform a laboratory experiment, which I present in the sixth chapter.

Kamien and Schwartz (1976) start with the hypothesis that there is an "optimal" degree of rivalry "between the extremes of monopoly and perfect competition". They draw this argument from several empirical studies. Scherer (1967b), using data on 56 industry groups, concludes that when the four largest firms in an industry exceed 50-55% of the market share, "additional market power is probably not conducive to more vigorous technological efforts". Mansfield (1963) investigates innovative behavior (as proportion of innovations introduced by the four largest firms) in the iron and steel industries during 1919-1938 and 1939-1958 and suggests that a lower concentration of market power would result in more innovation. His data is precisely quantified by Williamson (1965), who comes up with even more drastic recommendations: the four largest firms in an industry should have a combined market share of only 5-30% for optimal innovative activity (here innovative activity is defined as "proportion of innovations in relation to market share" (idem) of the largest players in the industry). Thus, Kamien and Schwartz (1976) develop a model to test the empirical finding that the rate of innovative activity increases with industry competition only up to a maximum point, after which it declines with further increase in competition. They derive the optimal level of "technological rivalry for innovation […] via comparative statics analysis of two related models of the profit-maximizing firm" (idem): in one model, imitators can obtain some rewards, while in the other model perfect patent protection is granted. Similarly to Kamien and Schwartz (1972), a firm chooses its own development period (introduction date), and there is the same assumption that the speed of innovation increases the costs. However, Kamien and Schwartz (1976) no longer have the assumption of uncertainty of development present in their paper from 1972. In their 1976 paper, Kamien and Schwartz assume that the development function is known with "a high degree of certainty", and it depends on "accumulation of effective development effort by expenditure of money". They find that stronger competition in an industry will first elicit a higher R&D investment by the firm, but will eventually cause the intensity of innovative activity to decline. This way, Kamien and Schwartz (1976) prove that there is generally some intermediate degree of rivalry, at which a firm's R&D expenditure will be maximal.

Loury (1979) questions the same argument studied by Kamien and Schwartz (1976) and the empirical studies associated with the later study: is there an optimal degree of competition (somewhere between pure monopoly and atomistic competition) that maximizes R&D performance? Loury (1979) criticizes the studies by Kamien and Schwartz (1972, 1976) on the grounds that the latter only perform partial-equilibrium analysis, “studying the behavior of an individual firm that views market conditions parametrically” (Loury, 1979). Loury points out that, in reality, each firm is the rival of all the other firms: "Thus, the likelihood of rival precedence depends on the R&D strategies chosen by other market participants, and cannot be treated as a parameter when analyzing changes in those decisions" (idem). Loury (1979) does his analysis under different assumptions than Kamien and Schwartz (1972, 1976). Loury’s assumptions are the following: a) the choice variable of a firm is R&D expenditure in a point in time; b) there is technological uncertainty for the development of the innovation (no firm can be sure when any of its rivals' R&D efforts will be successful); c) the probability of success is an exponential function of a hazard rate (a conditional probability density function expressing the idea that the present R&D efforts will affect the chance of R&D success in the next moment, given that the innovation has not succeeded yet); d) the completion time is an inverse of the hazard rate; e) investment is a sunk cost; f) equilibrium occurs when each firm's investment decision maximizes its profits, subject to the other firms' R&D investment strategies being given; g) rivalry is greater when the number of identical firms increases; h) the returns to R&D efforts initially increase, but eventually decrease (e.g. because the supply of successful ideas or talented researchers is limited). Loury's (1979) results are that an increase in competition will reduce firms’ individual investments, because competition diminishes the probability of success (and therefore the expected benefits). However, even though on an individual level investment decreases, through the larger number of firms present on the market, the aggregate level of investments will increase, and this will result in faster innovation. So in Loury's model, “atomistic competition is the market structure giving optimal innovative activity” (idem).

Dasgupta and Stiglitz (1980) also analyze the relationship between market structure and innovative activity. First of all, they point out to the many difficulties of measuring this relationship: that it is not the case of a single firm making an R&D expenditure decision, but several firms making “a complex of decisions"; that imitation behavior can lead to a wasteful duplication of R&D efforts; that excessively risky projects can lead to too high industry concentration and finally that each firm decision must be "made within an industrial structure which is itself endogenous" (idem). Dasgupta and Stiglitz criticize the general presumption (endorsed by Arrow's 1962 research) that there is under-investment in R&D both under competition, and under monopoly (this presumption is partly derived from the fact that knowledge has the characteristics of a public good). Dasgupta and Stiglitz underline that one of their most important conclusions of their 1980 paper is that, unlike in Arrow (1962), there can also be over-investment in the industry. Their model assumptions are: a) firms have fixed costs of R&D; b) firms choose quantities (through output and R&D expenditures), therefore there is Cournot competition; c) there is a perfect capital market, such that firms are not forced to resort to internal financing for the development of an R&D project; d) firms are engaged in a game of perfect information. Dasgupta and Stiglitz (1980) then solve each firm's maximization problem of the expected payoff with respect to the number of firms in the industry. Their results are quite complex. First of all, a pure monopolist has "insufficient incentive […] to undertake R&D expenditure" or "to engage in risky research ventures" (idem). On the other hand, competitive markets will "encourage firms to engage in overly risky research projects", and to excessively increase the speed of research (idem). However, "industry-wide R&D effort is positively correlated with concentration" in the industry (idem). Dasgupta and Stiglitz also show that their results become more ambiguous when we consider barriers to entry, the degree of risk aversion in society, and the elasticity of demand. My own conclusions from this study are that, except for the case of pure monopoly, a concentrated industry (with a few number of players) will lead to R&D efforts that are not too high and also not too low.

Lee and Wilde (1980) also compare a regime of perfect patent protection with a regime of competition. Lee and Wilde (1980) run a variation of Loury's (1979) model with one important modified assumption: a) instead of a fixed sunk cost (like in Loury, 1979), a flow cost (a variable cost) is paid, until the moment when some firm develops the innovation and wins the race. Further assumptions are: b) the timing of success depends on the size of this flow cost; c) competition should reduce both expected benefits and expected costs, because the flow investment will be made for a stochastically shorter period of time; d) competition is measured, like in Loury, as the total number of rivals in the industry. The results by Lee and Wilde (1980) are in total contradiction to Loury's (1979) results: the individual equilibrium investment will not decrease, but increase with competition! As a further result, aggregate investment in the industry will also increase with competition, and therefore, the completion time of the innovation decreases. Even more, opposing Loury (1979), an increase in aggregate investments (by all firms in the industry) will increase individual investments. Lee and Wilde's different results to Loury's study can be attributed to the modified assumption of the cost structure (which is a flow or variable cost in Lee and Wilde, and not a fixed cost as in Loury). However, the reliability of Lee and Wilde's results becomes questionable when one takes into account parallel research by other scholars. For instance, Futia (1980), who modifies the condition of perfect patent protection to allow imitation, obtains opposite results to Lee and Wilde: when imitation is fast and complete, there may be an inverse relationship between the aggregate investment in the industry and the number of firms (Futia, 1980). Moreover, Reinganum (1984) points out that one alternative form of imitation (different than the elimination of patent protection) is represented by positive knowledge spillovers. In the case of knowledge spillovers, aggregate investment will be inversely related to the number of firms (Spence, 1982), and results become similar to the case of regular imitation through the loosening of patent protection.

Reinganum (1982) develops a theory of dynamic optimal resource allocation to R&D in an industry with *n* firms. Reinganum emphasizes that she combines two alternative approaches which had been used in the literature to analyze R&D with competition: a static game theoretic analysis (Loury, 1979, Lee and Wilde, 1980, Dasgupta and Stiglitz, 1980) and a decision theoretic analysis (Kamien and Schwartz, 1972). In Reinganum (1982) the combination of the two approaches results in the following setting: rivals will be strategic agents, but their strategies will be functions of development time. The assumptions of Reinganum's (1982) model are: a) each firm has private knowledge about its own research; b) knowledge acquisition is achieved by the expenditure of resources, and it is deterministic; c) the success of the innovation is non-deterministic – the completion of the innovation is influenced, but not determined, by knowledge acquisition; d) the probability of success is an exponential function of the accumulated knowledge; e) research investment is a sunk cost; f) there are *n* firms; g) there is discounting. Furthermore, Reinganum compares one setting with perfect patent protection with one setting where imitation is allowed. Reinganum (1982) performs a dynamic game theoretic analysis by computing the Nash equilibrium in investment rules, where the optimal strategy depends on the number of firms, the innovator's and imitator's profit, time and the discount rate. Reinganum's (1982) results are that, first of all, perfect patent protection unambiguously increases the speed of innovation. This result is due to the increase in the rate of investment and the corresponding cost increase with time; as a consequence, knowledge generation and the probability of success also increase with time. Even an increasing number of rivals (higher competition) under perfect patent protection will lead to higher R&D investments, which also reduce development time. Secondly, Reinganum (1982) also analyzes the effect of imitation and concludes that it will depend on the ratio between the innovator's payoff and the imitator's payoff: the higher the appropriability of returns for the innovator, the higher will be investments into knowledge, and the faster the speed of innovation.

Reinganum's (1982) result that competition will lead to higher innovation speed is contradictory to Kamien and Schwartz (1976), who had shown that increasing rivalry will reduce R&D efforts after a certain point. Reinganum (1982) attributes this contradiction to the game theoretic analysis used in her paper *vis-à-vis* the decision theoretic analysis of Kamien and Schwartz (1976). Reinganum's (1982) results become even more ambiguous with imperfect patent protection, or competition: similarly to Kamien and Schwartz (1972), the existence of intense competition will make technical advance not a worthwhile option for the firms, so they may not embark in the development at all.

Reinganum's (1984) article justifies the use of game theory with R&D races, provides a literature review on the current research and suggests for the first time that patent race models should be empirically tested. She reiterates the argument that, depending on the cost structure (fixed cost in Loury, 1979 versus flow cost in Lee and Wilde, 1980) one obtains contradictory results: R&D investments decrease with competition for the case of fixed cost, but increase with competition in a setting with flow (variable) cost, and an increase in aggregate investment will reduce individual investment in the case of a fixed cost, but increase individual investment for the flow cost setting.

Furthermore, Reinganum (1984) points out that “there is an inverse relationship between the magnitude of an innovation and the likelihood that it is invented by a current industry leader”, independent of the cost type (fixed or flow). This is because an industry leader with a current advantage over the other firms has an incentive to protect his current profits, not to innovate. An innovation with technological uncertainty would bring the leader lower expected future profits. This argument is endorsed in the study by Freeman (1982), who shows that for large innovations market challengers will invest more in R&D than incumbent monopolists. A study by Ungern-Sternberg (1980) brings a complementary result to Freeman (1982) by proving that for small innovations the opposite is true: an incumbent monopolist typically invests more than a challenger, and the probability that the monopolist will succeed first decreases with the magnitude of the innovation. An individual firm's investment will decline with the number of anticipated innovations, because anticipated future innovations reduce the value of incumbency and increase the potential value to challengers. If the innovation production function is deterministic, the current industry leader will persist as a monopolist, because it will patent innovations before potential entrants do.

Last but not least, Reinganum's (1984) paper deals with licensing. Her main result is that licensing will encourage research when production costs are similar, but will discourage research when production costs are dissimilar (by taking a look at Fershtman and Markovich's (2010) study from Chapter 2, I found that Reinganum's (1984) result is contradicted and research is not discouraged for asymmetric costs). Reinganum's (1984) result on licensing is based on two different incentives to license: with similar costs, ex post incentives are large - firms are interested in monopoly and the reduction of production inefficiencies, which leads to more investment; with dissimilar costs, ex ante incentives are large - firms are afraid of the threat of low-cost competition and would like to eliminate any wasteful R&D expenditures, which leads to less investment. One solution proposed by Reinganum (1984) is licensing at a low royalty rate; this makes R&D expenditures unattractive to potential entrants, and subsequently the chance that challengers discover a very low-cost technology is eliminated or greatly reduced. A very important result by Reinganum (1984), which helps to create a bridge to asymmetric patent races (the topic of the next section) is that over the course of developing an innovation some firms fall behind and drop out of the research activity. The industry activity will become more concentrated. As we shall see, this result is confirmed for other asymmetric patent race models, such as Fudenberg et al. (1983) and Harris and Vickers (1985).

5.1.2. The 1980s: A Shift Towards Multistage Two-Player Patent Races

Most racing models from the 1970s and early 1980s, most notably Kamien and Schwartz (1972), Loury (1979), Lee and Wilde (1980), Dasgupta and Stiglitz (1980) and Reinganum (1982) assume that there is a single stage of development in the innovation process, and all competing firms have the same technological potential. These models are static and symmetric - in the sense that firms are considered identical and competition is finished when any of the firms completes the single development stage. The models are based on quite unrealistic assumptions, because most innovations require several stages of development, and firms are usually in asymmetric positions - that is, technologically ahead or behind. Moreover, learning is not possible in the mentioned models. Due to these shortcomings, many researchers felt the need to develop alternative patent race models in order to accommodate dynamic aspects of R&D competition, asymmetries between industry players and learning effects.

In the following, I describe the most important of the theoretical models on asymmetric races with leader and follower. Few of these models were considered suitable for experimental testing; my experimental investigation from Chapter 6 is also based on an asymmetric setting. For theoretical models with experimental counterparts, a brief comparison of their results is provided in this section.

Fudenberg et al. (1983) develop a model of a dynamic multistage stochastic innovation process. The model was experimentally tested in the study by Silipo (2005), which I describe in Chapter 5.2. The main conclusions of Fudenberg et al. are overthrown by Silipo's experiment (however, we shall see that Silipo significantly altered Fudenberg et al.'s assumptions). The paper by Fudenberg et al. (1983) investigates whether the patent race will be characterized by strong competition, or whether it will converge into a monopoly. Fudenberg et al. identify three cases: 1) there is ε-preemption, that is, one firm has an “arbitrarily small headstart” over its rival, innovation is deterministic, and the rival has no chance of catching up; 2) the multi-stage game allows for "leapfrogging" in the sense that the large number of steps necessary to complete the innovation permits the rival to compete and jump ahead of the other firm; 3) the patent race is deterministic but there is lagged information, the monitoring of R&D efforts is imperfect, firms cannot be fooled by empty threats and for these reasons both players may compete. The first case of ε-preemption is present in the model by Dasgupta and Stiglitz (1980) which I discussed before. In Fudenberg's adaptation a firm's current chances of making a discovery depend only on its “stock of experience”, where experience is a function of R&D engagement. ε-preemption is induced by the fact that R&D is a viable activity for a monopolist, it is not profitable for both firms to engage in it, and one firm enters the patent race before the other firm. Fudenberg et al. compute the equilibrium, in which one firm engages in R&D, and the other firm drops out of the patent race. Cases 2) and 3) represent situations of vigorous, continuous competition, the basis of other models such as Loury (1979), Lee and Wilde (1980) and Reinganum (1982). In these models, preemption is not possible and firms compete until a discovery is made. In case 2) R&D takes place in two stages. In the first stage there is preliminary invention (e.g. the concept of the research program is defined), while in the second stage there are several R&D progress steps (Fudenberg et al., 1983). Just like in case 1), one firm begins the race before the other firm, and there is no room on the market for both firms (because they would make a loss). The probability of success at a certain point in time again depends on experience, but there is a chance for the lagging firm to complete the first stage before the leading firm, resulting in the leading firm dropping out before the second stage begins. This phenomenon is defined as “leapfrogging” (idem). In case 3) modeled by Fudenberg et al., a firm may make progress without revealing information to another competitor, or information is lagged in the sense that a firm at time t only knows about the R&D activities of competitors up to time $t - 1$. In each point of time, a firm may invest zero effort (yielding no additional knowledge), low effort (leading to one unit of knowledge), or high effort (resulting in two units of knowledge). Discovery occurs when any firm accumulates *N* knowledge steps. The authors show that in any perfect equilibrium, if the lag between the firms is equal to or larger than two steps, the follower drops out. If the lag equals one step, the follower decides between high effort and quitting, while the leader decides between high effort and low effort. If the lag is zero (firms are tied), they compete vigorously if the number of steps left is small enough, and otherwise they randomize between low and high effort. The main departure of the model by Fudenberg et al. (1983) from the models by Loury (1979), Dasgupta and Stiglitz (1980) and Reinganum (1982) is the fact that experience does not play a role in the latter models. In Fudenberg's model, experience is essential and it might cause the patent race to degenerate to a monopoly from the start (in the extreme case 1), while in the multi-stage cases 2) and 3) strong competition is possible.

Harris and Vickers (1985) also develop an asymmetric model of a multistage stochastic innovation process of a patent race and they explain the reasons why this asymmetry might exist. I describe these reasons, closely following the authors' argumentation. For once, the two firms could value the patent differently: the incumbent, who wants to keep her monopoly position, is more probable to value the patent higher than a challenger. Or, the incumbent has more knowledge due to her market experience. The third possibility is that firms have different R&D efficiency levels. In their paper, Harris and Vickers (1985) focus on two cases: one "pure" case, in which the challenger's technology is not superior to that of the incumbent (here the incumbent's only concern is to prevent the rival from winning the patent) and one "hybrid" case, in which the new technology is superior to that of the incumbent (here the incumbent has an additional incentive to win the patent, in addition to stopping the rival). The main result of Harris and Vickers (1985) is that the incumbent has considerable advantages from her asymmetric position, and she will also go to great lengths in order to keep her advantage. This result is also confirmed in the more advanced model by Harris and Vickers (1987), which I describe in the following.

Harris and Vickers (1987) is another theoretical patent race model which was later experimentally tested in the prestigious publication by Zizzo (2002), which I discuss in Chapter 5.2. Zizzo's experiment found only limited support for the conclusions of Harris and Vickers, while the most important of the theoretical study's results were downright contradicted by the experimental test. In the model by Harris and Vickers (1987), two firms compete for the multi-stage development of an innovation with the goal of achieving patent protection. In each round their strategy choice is the investment effort rate, which is x for firm 1 and y for firm 2. In turn, the investment effort influences the probabilities of winning in any round. These probabilities are $\frac{x}{x+y}$ for firm 1 and $\frac{y}{x+y}$ for firm 2. Whoever wins a certain number of rounds completes the innovation, wins the race and gets the patent. The main assumption about costs is that the cost function is convex in investment: $c(x) = x^{\mu}$. In the model, a quadratic cost function was mostly used: $c(x) = x^2$. Harris and Vickers explain in mathematical terms that the convexity of the cost function is a necessary condition for the existence of a unique equilibrium in their model. However, in this dissertation I regard the convexity assumption as questionable. Cost convexity implies that it is easy and cheap to develop the first units of knowledge, but increasingly expensive to generate later knowledge. In my view, reality could be different: setting up a research facility generates large sunk costs and little knowledge in the beginning; at later stages of development, learning effects could increase the ratio of knowledge to costs, leading to a concave cost function. The main results of the study by Harris and Vickers (1987) are that leaders make greater investment efforts than followers, and the leader invests more the larger the gap relative to the follower. The follower invests more the smaller the gap relative to the leader, so the correlation between gap and investment should be more negative for followers than for leaders. When the gap becomes large enough, the follower drops out completely and the leader does almost all of the investment by herself. Therefore, investments are greatest when the gap between competitors decreases. The latter result is in tune with the predictions by Grossman and Shapiro (1987), who find that players close to each other invest more when the gap is small (e.g. one step) than when the gap gets larger (e.g. two steps or more). That is, if players are neck-to-neck in the race and they have fewer progress steps to go, they should invest more than if they have more progress steps to go.

Grossman and Shapiro (1987) analyze a dynamic R&D model and look at how the effort level by firms varies over the course of a race. Like in most previous patent race models, success is uncertain, the probability of breakthrough depends on the allocation of resources, and there is a "winner-take-all" setting, so that inventing around the patent is not possible. Grossman and Shapiro criticize the model by Fudenberg et al. (1983), which according to the former authors severely restricts a firm's ability to vary its R&D intensity. Grossman and Shapiro build on the model by Lee and Wilde (1980) by adopting their stochastic structure. However, Grossman and Shapiro "allow for progress […] by introducing a single intermediate step in the research programme facing each firm". The research program requires the completion of two phases of equal difficulty, where "the two stages may be thought of as research and development, respectively" (idem). Another assumption of the model is that each firm has full information about the progress of its rival. Grossman and Shapiro state that, even with this simple structure, a "rich set of possibilities" is encountered, depending on the parameters of the model. In the results section, competition is shown to be "most intense when both firms are even and each has completed the initial phase of the research project. When a lagging firm draws even with a rival that was formerly ahead in the race, both competitors respond by increasing their research efforts. […] When the two firms are at different stages in the innovation process, the one that is ahead has a greater incentive to invest in R&D than the one that is behind" (idem). These results confirm the theoretical predictions by Harris and Vickers (1987) as well as Fudenberg et al. (1983).

Park (1987) suggests in his paper that he "adopted a simple structure" for his racing model, but keeping some "essentials of innovation competition such as stochastic R&D production, multiple stages, and choices with varying degrees of risk". Park's model is based on Lee and Wilde's (1980) single stage model of competition, which was extended to a two-stage model of development. Park (1987) says that the main objective of his paper is to see "how the expected benefits, the cost of R&D, and interactions between competing firms combine to determine their dynamic R&D strategies over time". More precisely, Park analyzes how a firm changes its R&D strategies depending on its competitive position (ahead or behind). There are two possible development paths leading to the final innovation: one is "a big simple jump with a small success rate" ("all-or-nothing" strategy), the other is "a stage-by-stage development" which progresses slow but safe (idem). In the paper, the cost of the two development paths is considered to be equal, an assumption I shall use later in the experimental chapter of this dissertation. Park's results are again similar to those of Fudenberg et al. (1983), as well as Harris and Vickers (1987) and Grossman and Shapiro (1987): a leader allocates more resources to R&D than a follower, because the value of an R&D race is greater for the leader than for the follower. In Park's model, the follower prefers the riskier big jump strategy to the safer stage-by-stage development and the leader sticks with the safe strategy. Moreover, competition is shown to be more intense when development is more advanced - that is, when the current stage is late rather than early.

Another branch of the patent race literature that is worth mentioning in the context of multistage (two-player) patent races is concerned with sequences of innovations where the technological history of the firms matters. Here I have to name the prominent papers by Reinganum (1985), Vickers (1986) and Beath et al. (1987), fulminating in the paper by Delbono (1989). The paper by Reinganum (1985) somewhat jumps out of my definition of two-player races, as she analyzes an *n*-player asymmetric model with one incumbent and several challengers. In Reinganum's (1985) paper, firms compete to introduce a sequence of innovations. The most important assumption she makes is that innovations are considered "drastic", meaning that whoever innovates obtains an advantage over the previous technology and can earn monopoly profits until the next innovation is introduced. Reinganum refers to this as the equivalent of the Schumpeterian process of "creative destruction" (Schumpeter, 1942). The main conclusion from Reinganum's paper is the following: creative destruction is due to the fact that the incumbent has an incentive to impede innovation and protect her current position in order to continue to earn monopoly profits. Her rival will therefore intervene at some point, inducing creative destruction. Beath et al. (1987) criticize Reinganum's (1985) "drasticness" assumption and show it may be correct for process innovation in fast-changing industries, but it does not hold for product innovation. With the "drasticness" assumption invalidated, Beath et al. (1987) point out that firms' “profits from winning or losing a particular patent race will depend on previous history and the sequential structure is therefore essential”. Beath et al. (1987) also pick at the fact that in Reinganum (1985) the sequential structure does not have an effect on the general conclusion from the paper (which was stated above). Beath et al. emphasize the central role of the sequential innovation structure: firms are interested to win a particular race “not just for the immediate profits it brings, but also for the strategic advantage/disadvantage it may confer in subsequent patent races". The same reasoning can be found in the paper by Bessen and Maskin (2009), who deal with sequential innovation, and in the paper by Vickers (1986) I mentioned just above. Vickers (1986) develops a model in which firms bid for patents in an auction, and the highest bid wins the patent. Vickers shows that if prices and quantities are determined in a Cournot equilibrium (generated by quantity competition), firms will engage in a game of action and reaction (or Catching Up as we shall call it below) with falling prices over time. If firms are involved in a Bertrand (price) competition, one firm will persistently dominate the other (which later is denoted as Increasing Dominance). Both the paper by Beath et al. (1987) and the later one by Delbono (1989) are extensions of Vickers (1986). Beath et al. are concerned to see how Vickers' insights carry over to sequential product innovation. Beath et al. point out to a shortcoming of Vickers' model: the latter only works under the assumption of slow technical progress. If, however, technical progress is sufficiently fast, the model by Vickers becomes indeterminate (Beath et al., 1987). The latter authors say that it is vital to distinguish between process innovation, in which case a firm "will only ever employ the single best technology for which it currently holds a patent", and product innovation where the same principle no longer applies (idem). Beath et al. find that "firms' responses are far more complex in the case of product innovation": here, firms may wish to respond by using a set of products rather than just one single product, and the consequence can be either action-reaction (Catching Up) or Increasing Dominance. With process innovation, firms respond with their latest or best technology and the consequence is only Increasing (or persistent) Dominance on the market. Delbono (1989) also builds on Vickers (1986) but reformulates the latter's model to also account for incremental innovations. In Delbono's paper, "payoffs – and thus profits and incentives – […] depend on the total number of patents previously won by the firms". Delbono underlines that this setting covers situations in which technological changes "do not allow the innovator to overtake the rival" and thus depicts the case of incremental innovations. Delbono finds that with incremental innovation, "both Bertrand and Cournot competition imply Increasing Dominance". What we can see is that even in a small niche of the patent race literature containing the four papers described in this paragraph, researchers have dealt with quite specific assumptions and heterogeneous settings. Consequently, their results are just as specific, heterogeneous and impossible to generalize.

5.1.3. Reinganum's (1989) "Excellent Review"

The review of patent race literature done by Reinganum (1989) was highly praised in the scientific community, with virtually every subsequent patent race paper quoting the review. In this section I sketch her most valuable findings.

Reinganum (1989) classifies the pre-1989 patent race literature into four streams:

a) symmetric noncooperative models; most notable here are Scherer (1967a), Dasgupta and Stiglitz (1980), Loury (1979), Lee and Wilde (1980), Reinganum (1982), Kamien and Schwartz (1972), which I have dealt with earlier in this chapter and have largely called the "traditional" patent race literature;
b) asymmetric models; among others I found Arrow (1962), Reinganum (1985), Fudenberg et al. (1983), Harris and Vickers (1985), Grossman and Shapiro (1987). Additionally I covered papers which are not included in Reinganum (1989): the important paper by Harris and Vickers (1987), which had an experimental implementation, as well as the papers by Park (1987), Vickers (1986), Beath et al. (1987) and Delbono (1989);
c) models of licensing; I touched the topic of licensing as I looked at the studies by Reinganum (1984) and Fershtman and Markovich (2010). A thorough analysis of licensing models within patent races is beyond the scope of this dissertation;
d) models of innovation adoption and diffusion; these models are also not dealt with in the dissertation.

The review by Reinganum (1989) also emphasizes the existence of two paradigms which may come into play within every of the scientific papers listed above: the first paradigm is that of deterministic innovation, while the second paradigm involves stochastic innovation. Under the two different paradigms, results of the literature may vary considerably (as we can see below in this section). Moreover, Reinganum (1989) formulates 16 assumptions and 40 propositions which capture the mathematical specifications of her selected papers. I limit my analysis to a verbal explanation of her findings.

Reinganum’s (1989) main results from the first literature stream of symmetric noncooperative models are as follows: aggregate expenditure on R&D in an industry is “too high relative to the cooperative optimum”; there is overentry and overinvestment, until all profits are dissipated; firms “forego intertemporal efficiencies which could be realized by investing at a lower rate over a longer time planning horizon” (idem).

The second literature stream including asymmetric models is especially relevant to this dissertation due to the already performed experiments (albeit with results in non-concordance with theory). I adopt the asymmetric setting as the most interesting case in Chapter 6. Reinganum (1989) points out that results under the asymmetric setting are “particularly sensitive” to the paradigms mentioned above: “the presence or absence of technological uncertainty” (idem). If the innovation is uncertain, the leader invests less than the follower, and the role of the leader changes from one firm to another (in Reinganum’s words, it “tends to circulate around the industry”). This diminishes the value of future races for the leader, because firms expect the leader to change constantly; at the same time it increases the value of future races for the follower, because she can still win the race through luck (in technical terms, luck is represented through the stochastic nature of the innovation). If, differently, innovation is deterministic, Reinganum shows that in most models the leader invests more than the follower. The role of the leader stays constant and persistent. This increases the value of future races for the leader, because she then knows she can keep her position forever. My experiment in Chapter 6 has a very special nature: I model the race so as to incorporate both a setting of uncertain innovation, and a setting of deterministic innovation. I therefore experimentally test the very core of the asymmetric patent race literature.

It is worth mentioning Reinganum’s (1989) results from the third and fourth streams of literature named above, which are based upon the analysis of around fifteen scientific papers. For licensing models, Reinganum finds that research joint ventures have fewer incentives to innovate and also reduce the amount of innovation diffusion. However, there are some benefits: scale efficiencies occur and the wasteful duplication of R&D efforts criticized in the literature is avoided. In models of innovation adoption and diffusion, Reinganum finds the following: without network effects, each firms invests too early, thereby reducing the value of adoption for the other firms. With network effects the opposite is true – there is "excess inertia", or no firm invests early; rather, each firm waits to imitate the innovation.

5.1.4. Further Ramifications of the Patent Race Literature

There are several ramifications of the post-1989 patent race literature which I have selected because they were published in highly ranking journals. Examples may range from Hartwick (1991), who links patent races to firms' market entry and bases his work on Lee and Wilde (1980), to Baye and Hoppe (2003), who show the strategic equivalence of innovation tournaments, rent-seeking contests and patent races. Such examples could surely continue, but they only bring additional complexity to an already hardly comprehensible literature. For the sake of having delivered a nearly complete image of the patent race literature, this section will briefly present two further ramifications of the literature: perpetual (or endless) races and optimal rules for patent races. I have chosen these ramifications because perpetual races are linked to an experimental investigation, while the other ramification is based on a very recent study and tries to synthesize an optimal design for patent races, which I regard as an appropriate end for Chapter 5.1.

Perpetual races were first analyzed in the paper by Aoki (1991), while Hörner (2004) builds up closely on Aoki by also modeling such perpetual races. Looking further, one finds that the theoretical model by Hörner was tested experimentally by Breitmoser et al. (2010), a paper I describe in Chapter 5.2.1. Interestingly, the results by Breitmoser et al. support the reasoning I formulate in my experimental Chapter 6. Breitmoser et al. point out, just as I do, that the predictions of theoretical models are highly sensitive to small changes in the parameters; however, when put in an experimental setting, subjects' strategies are far less sensitive. With patent races, people are simpler than models. In the case of perpetual races, Breitmoser et al.'s experimental results are more stable than the predictions of the theoretical models by Aoki and Hörner.

Turning back to the paper by Aoki (1991), we find a model of two firms who are equally efficient and engage in competition for R&D projects in an infinite number of periods. In Aoki's basic model, the R&D technology of the firms is assumed to be deterministic, with any firm choosing between: i) making a costly effort and advancing by one unit of knowledge and ii) making no effort and staying where it is. Only the firm with more units of knowledge may sell the product and may earn a monopoly profit. Aoki computes the resulting equilibrium: for a sufficiently low cost-profit ratio, both firms invest when they are in even positions, the follower is indifferent between decisions i) and ii) when she is one step behind the leader, and the follower drops out if the gap is equal to or exceeds two, because in this case the follower's continuation value becomes zero. So the prediction of Aoki's model is that investments in R&D will be largest and competition fiercest when firms are closest to each other.

Hörner (2004) draws on the models by Aoki (1991) and by Harris and Vickers (1987). Hörner uses Aoki's assumption that the leader of the race earns a larger payoff, as well as Harris and Vickers' setting that firms may decide between exerting high and low effort, where high effort is more expensive than low effort but high effort increases the probability of making technological progress. Similarly to Aoki, Hörner models a "dynamic competition between two firms that repeatedly engage in an innovative activity" with an indeterminate time horizon (idem). Differently from Aoki's deterministic innovation setting, in Hörner the "state of competition" (defined as "the difference between the number of innovations introduced by the firms") "evolves stochastically according to their effort level" (idem). As we already have gotten used in the world of theoretical patent races, Hörner's results contradict Aoki's, probably due to the stochastic and not deterministic nature of the innovation. Hörner finds it generally not true that competition and therefore effort levels are highest when firms are closest, but rather competition is strongest when a leader is sufficiently ahead, as she tries to secure her position. Competition is also high when the lead is very small, as firms try to defend or regain leadership; when the gap between firms is of intermediate length, firms' effort levels will be lower (idem).

The last paper I present in this chapter is the one by Judd et al. (2012), who derive optimal rules for patent races. Judd et al. emphasize the fact that their paper lies at the intersection of the literature on patent regulation and the literature on patent races; they combine elements from both streams of literature in order to find an optimal patent race design. According to Judd et al., a patent granting authority (the regulator) can influence the outcome and optimality of a patent race by using two instruments of regulation, which I describe closely following Judd et al.'s reasoning:

- the race length, which is determined by the minimal accomplishment required from the innovating firm(s) in order to be granted a patent. A short race (or early grant of the patent to one firm, after only few steps of the innovation development) has the disadvantage that it gives an extended monopoly to the patent holder. On the other hand, a long race (or late grant of the patent) induces many firms to get involved in parallel developments of the same innovation and leads to a wasteful duplication of development costs;
- the size of the prize, defined as the benefits accruing to the innovating firm from receiving of a patent. The regulator may increase the size of the prize by: granting long patents, granting broad patents or charging small patent renewal fees. The regulator may choose between two strategies: "offer a big prize to a single innovator or offer a smaller prize but use a race to threaten each firm that the prize will go to its competitor" (idem).

Judd et al. (2012) continue by showing that the regulation of these two instruments makes the patent granting authority face a series of difficult trade-offs. For one thing, the regulator would like to speed up the innovating process because this benefits consumers. However, an early introduction time of the innovation by a firm comes with some disadvantages: over-investment cost and welfare loss through monopoly. The over-investment cost, or the waste, can be reduced by setting a short race (early patent grant), but this makes it "difficult to filter out the less efficient firm", because even firms with high development costs are motivated to join the race for only a few stages (idem). Both the over-investment cost and the welfare loss through monopoly can be reduced if the regulator decreases the reward (the prize) by diminishing the patent length or breadth. But then firms have fewer incentives to innovate fast (because this is more costly), which brings the regulator back to the problem she wanted to solve in the first place (faster innovation).

To arrive at recommendations about an optimal race design, Judd et al. develop a dynamic multistage innovation race model between two firms with heterogeneous cost of R&D effort. Their model is very complex and contains a variety of parameters: the time of granting the patent; the time of innovation completion; the social value of the innovation; the prize to the innovator as a fraction of the social value; the ratio between deadweight loss and monopoly profits. Furthermore, costs may be linear or convex, and the analysis is done under two settings: when the regulator is interested to maximize the social surplus, or to maximize the consumer surplus. Judd et al. solve their model by performing long and complicated numerical analyses and provide their results, which I synthesize in the following.

Under most circumstances, races of "nontrivial duration" are part of an optimal policy, because they: a) motivate "firms to invest and complete the innovation process quickly" and b) help filter out inferior innovators with high development costs (idem). The decision between short and long races depends, according to Judd et al., on the regulator's objective (social versus consumer surplus maximization), on the "social returns to innovation" and on the "inefficiency costs of compensating the patent winner" (idem). A patent granting authority who wants to maximize social surplus should choose races of medium length and small prizes in case deadweight losses are present. If deadweight losses are excluded from the analysis, optimal policy becomes conditional on the firms' cost asymmetry: when firms are homogeneous in terms of costs, large prizes and no races should be chosen; when firms are heterogeneous in their costs, long races and smaller prizes are socially preferable. If the patent authority wants instead to maximize consumer surplus, it should choose lower prizes and grant the patent at later stages, as these strategies transfer benefits directly from the innovators to consumers. Another important contribution by Judd et al. (2012) is that they present two quite viable alternatives to patent races. One alternative is to finance the prize through taxation and give the prize to the inventor, but then to make sure the innovation is produced under a regime of competition. This would "avoid the monopoly inefficiencies of a patent but at the cost of distortionary taxation to finance the prize" (idem). The second alternative is represented by research tournaments, in which "contestants compete to find the innovation with the highest value to the sponsor" (idem). Judd et al. point out that research tournaments are "particularly useful when research inputs are unobservable and research outcomes cannot be verified by courts". In an innovation race "the quality requirement is fixed" and "the time of innovation is variable", while in a tournament "quality is variable" and the terminal date is fixed (idem).

5.1.5. The Theoretical Patent Race Literature Is Inconclusive

Throughout this chapter we have navigated through a very stormy sea of theoretical patent race models – a literature full of contradictions and lacking stability and reliability. Early in this literature, Scherer (1967a) as well as Kamien and Schwartz (1972) were signaling that, in models of competition with R&D, small changes in parameters will have unpredictable consequences on all model results. We have seen in this chapter that virtually any of the theoretical papers on patent races published in highly ranking journals after Barzel (1968) only added to the complexity and confusion, without bringing clarifying elements. Kamien and Schwartz's (1972) results contradict those by Barzel (1968). Lee and Wilde (1980) contradict Loury (1979). Reinganum (1982) contradicts Kamien and Schwartz (1976). Compatible, harmonious results are found in the papers by Fudenberg et al. (1983), Harris and Vickers (1985), Park (1987), Grossman and Shapiro (1987) as well as Harris and Vickers (1987). But the results by Fudenberg et al. (1983) are overturned in the experimental study of Silipo (2005) and Harris and Vickers (1987) are mostly contradicted by the experimental findings of Zizzo (2002). The harmonious results I mentioned fail to withstand experimental testing. The ramifications of the patent race literature I discussed in this chapter - the models where technological history matters, the endless race models and the paper on optimal rules for patent races - address research questions that are incompatible and only increase the literature's complexity. Reinganum's (1989) excellent review does a good job at categorizing the different models and it provides valuable general conclusions for each stream of the pre-1989 literature. Reinganum (1989) successfully synthesizes the results of this literature into 40 mathematical propositions and 16 assumptions, which are of great theoretical value. However, a subsequent study by Cockburn and Henderson (1994) found it hard to match these theoretical results with empirical evidence, as it is difficult to find features of industries to which the theoretical findings can be applied.

Cockburn and Henderson (1994), who try to test some assumptions and results of the patent race literature in the pharmaceutical industry (in which patents and racing are considered to be of outmost importance), argue that in general patent race models are "very difficult to test". The authors point out that firms' investment behavior "is critically dependent on a wide range of factors, including the nature of the payoff function, the spillover regime, the information structure of the game and the extent and nature of asymmetries between players, all of which are rather difficult to capture empirically". Cockburn and Henderson also assert that only "few of the models in the existing literature can be easily nested", that "discriminating empirically amongst these models is a daunting prospect" because theoretical models trying to simultaneously include interactions between the relevant factors "quickly become intractable", and that conclusions from the models are fragile, with small changes in any factor being able to "easily reverse or weaken any given theoretical result".

These considerations are in tune with the high degree of heterogeneity between results of studies which we encountered in the literature review of this chapter. The theoretical patent race literature is too complex and unable to deliver any reliable predictions. Each theoretical paper contradicts some other one, or studies aspects (parameters, settings, research questions) that are incompatible with other papers. Most importantly, theoretical models also need experimental testing. As we shall see in the next chapter, the theory from the patent race literature is largely contradicted by experiments.

5.2. How the Experimental Evidence Contradicts the Patent Race Literature

This chapter is separated into two subchapters. In Chapter 5.2.1 I analyze the four experimental studies which took models from the theoretical patent race literature as their basis, and I compare the results from theory with those from experiments. In Chapter 5.2.2 I describe patent race experiments which, differently from those in 5.2.1, did not root their theoretical basis in the standard patent race literature, but instead developed their own theoretical models and then tested them experimentally. Towards the end of both subchapters, I provide summarizing tables with the features of each experimental study.

5.2.1. Experiments Testing the Theoretical Patent Race Literature

As we have already seen in the previous chapter, Silipo's (2005) experimental study is based on the theoretical model by Fudenberg et al. (1983); Harris and Vickers' (1987) study, a theoretical extension of Fudenberg et al., was tested in Zizzo (2002) and Kähkönen (2005); Hörner's (2004) study, a theoretical extension of Aoki (1991) and Harris and Vickers (1987), was tested experimentally by Breitmoser et al. (2010).

Silipo (2005) experimentally investigates cooperative versus competitive behavior with respect to the relative position (gap) between two subjects (firms) in a patent race setting. Even though Silipo's study is based on the theoretical model by Fudenberg et al. (1983), it contains a significant departure from the latter: Silipo adds the possibility that the subjects make the R&D investment jointly, in order to save costs. A cooperation of this sort was not possible in Fudenberg et al., and therefore I have to underline from the beginning that the experimental study does not represent a direct test of the theoretical model.

The reason why Silipo (2005) introduces the option of cooperation between firms involved in a race can be well understood. Previously, Vonortas (1997) as well as Hagedoorn (2000) had found that both the creation and the falling apart of joint ventures represented essential factors in the development of R&D cooperation in the United States between 1985 and 1998, with a surge in the generation of research partnerships between 1985 and 1995 and some decline afterwards. On one hand, Silipo emphasizes the importance of cooperation and R&D joint ventures in reality, and on the other hand Silipo gives three reasons for choosing the model by Fudenberg et al. (1983) as a theoretical basis for his experiment. First of all, the model by Fudenberg et al. takes into account the learning process and the accumulation of knowledge during the patent race (which was absent in most of the other patent race models). Second, Fudenberg et al. allow firms to change their decisions during the race, depending on their relative position, and this allows us to analyze both situations: when they are symmetrically or asymmetrically positioned. Third, the model by Fudenberg et al. is more appropriate to examine the effects of the game's information structure on innovation and cooperation incentives. So it is easily understood that Silipo wants to get the best of both worlds: from theory and from practice. However, by introducing the option for cooperation Silipo (as he states himself) makes a radical departure from Fudenberg et al.'s model – an essentially different model is now being analyzed, which I regard as a very critical problem for Silipo's paper as well as for the consistency of the patent race literature.

In relation to the theoretical model, as in Fudenberg et al., in Silipo's paper R&D is deterministic and subjects participating in the multi-stage race make a costly effort to accumulate units of knowledge over time. The first subject to reach N units of knowledge wins the race and is awarded the patent. In relation to the experimental implementation, Silipo's experiment was run with 172 subjects who played 86 races in dyads. In each round, the players could accumulate 0, 1 or 2 units of knowledge with convex costs ($0, $0.44 and $1.11, respectively). Silipo designed his experiment such that no player could buy two units of knowledge more than ten times. That is, no player could accumulate more than 20 units of knowledge in large chunks (chunks of 2 units each), and after the achievement of 10 x 2 = 20 units, each player was forced to choose either 0 or 1 unit of knowledge per round. In other words, there is a maximum cap on the choice of the high effort strategy (in the form of 2 units of knowledge). I regard this issue as a very critical assumption and Silipo does not give any reason for the decision to design the experiment in this fashion. Each round consisted of two stages. In the first stage, subjects decided whether to cooperate or to invest alone. In the second stage, if cooperation had been agreed on, subjects undertook the R&D effort jointly at a cost advantage; if cooperation had not been reached, each subject made an independent and simultaneous investment decision. The winner of the race was the first player to achieve $N = 30$ progress steps, and her payoff equaled the prize minus the total costs. If both players cooperated throughout the race and reached 30 steps jointly, they split the difference between the prize and the total costs. Silipo's experiment had a between-subjects design with six treatments generated by varying the prize (high prize of $37.94; low prize of $18.97) and the initial positions of the two players (symmetric, both starting with (0,0) units of knowledge; asymmetric, with (0,1); asymmetric, with (0,2) units of knowledge).

Silipo's (2005) findings contradict Fudenberg et al. (1983) as well as Harris and Vickers (1987), who had predicted that firms would compete vigorously when close to each other (when the gap is small) and that the race would degenerate into a monopoly when the gap is large. As stated above, this should come as no surprise, because Silipo's introduction of the cooperative option changes the research question. Silipo's results are that subjects starting from the symmetric position (0,0) typically cooperate at the beginning of the race to save on investment costs; as they approach the finishing line, subjects break cooperation because of the incentive to get the monopoly prize. Subjects starting with asymmetric position (0,1) were found to cooperate if the cost saving is large enough to turn an unprofitable project into a profitable one. Subjects starting with asymmetry position (0,2) never cooperate. Silipo's main conclusion is that, if the cooperative option is included, subjects or firms would use it, leading to less wasteful duplication of investment efforts (which is quite an intuitive result). Furthermore, Silipo finds that in general firms cooperate when they are close to each other (opposite to Fudenberg et al., 1983 and Harris and Vickers, 1987), in order to reduce the costs of the innovation, but they break the cooperation as they come closer to the finishing line, because the cost-reduction effect weakens and the incentive to win the monopoly prize increases. Firms far away from each other never cooperate. A further result by Silipo is that cooperation reduced the speed of innovation when compared to a competitive R&D market.

Zizzo (2002) performs an experimental investigation of the multi-stage patent race model by Harris and Vickers (1987). Zizzo first of all emphasizes the importance of his research: it tests for dual uncertainty. Real-world patent races are confronted with two types of uncertainty: technological uncertainty (a given investment in R&D leads to an uncertain output in the development of an innovation) and dynamic uncertainty (firms' "incentives to invest in R&D may change as the race unfolds, according to the position of a firm in the race relative to its competitors and relative to the end of the race", idem). Zizzo points out that the earlier work on patent race theory (most notably Dasgupta and Stiglitz, 1980 as well as Gilbert and Newbery, 1982) focused only on the first type of uncertainty, the technological one. As emphasized by Zizzo, in these types of models the uncertainty is modelled as an exponential process and the patent race is "memoryless" in the sense that firms decide in a single round how much effort they want to put up, after which the winner or loser of the race is determined. Later papers (such as Fudenberg et al., 1983 or Harris and

Vickers, 1985) concentrated on the second type of uncertainty only, the dynamic uncertainty. Zizzo writes that the latter research "factored out technological uncertainty by having a deterministic relation between investment and progress in the race". The result of these models is the so-called ε-preemption: a small advantage on the side of one firm (the leader) at the beginning of the race ensures that the leader wins the entire race and the follower drops out immediately after the race begins.

Zizzo (2002) criticizes the above-mentioned papers because "neither lack of memory nor ε-preemption appear plausible features of real-world races". Instead, Zizzo prefers to experimentally test a theoretical model which reunites technological and dynamic uncertainty, as this would lead to "a better characterization of patent races" (idem). For this reason, Zizzo chooses the model by Harris and Vickers (1987) as a suitable benchmark (Zizzo points out that Grossman and Shapiro, 1987, who also deal with dual uncertainty, are just a special case of Harris and Vickers, 1987, so the latter paper contains the more general and thus more appropriate model). So Zizzo builds an experiment around Harris and Vickers (1987), trying to depart as little as possible from the theoretical model. Zizzo keeps the multi-stage nature of the race as well as the investment levels, probabilities of winning and cost considerations present in the theoretical model. The only difference to the theoretical model is the introduction of a budget constraint, which is necessary to run the experiment, for practical reasons (subjects cannot be endowed with infinite amounts of money). Zizzo tests in his experiment the main results of Harris and Vickers (1987): leaders should invest more than followers, and even more with a larger gap; the correlation between gap and investment should be more negative for the followers; when the gap gets large enough, the leader should do all of the investment.

In Zizzo's experimental design 36 subjects (nine sessions of four people each) played in dyads a "prize race" (with one practice stage, one payoff-relevant stage and changing counterparts between the two stages). In every race subjects were endowed with 500 points and the value of the prize was set at 1000 points. In every round of the race, the two players of a dyad had to decide on the levels of their investments x and y, where their probabilities of winning the round were $\frac{x}{x+y}$ and $\frac{y}{x+y}$, as in the model by Harris and Vickers (1987). The cost of an investment x was convex, assuming the quadratic form $c(x) = x^2$, and it was subtracted from a subject's endowment with each investment decision. As I stated before in Chapter 5.1.2., the convex cost function is in my view a hard assumption, because it assumes that the cost of developing the project grows exponentially with each new unit of knowledge. I believe that economies of scale and learning effects with R&D projects should lead to a reduction of the marginal development cost and therefore to a concave (not a convex) cost function. In Zizzo's experiment, if a subject won a round, she would make one progress step; if both players invested zero, no subject made any progress. After the decision was made, information on the winning probability, the investment of the other player, and the winner was provided. The first player to complete 10 progress steps became the winner of the race and received the prize of 1000 points. Again, the only structural difference to Harris and Vickers' (1987) model was that subjects faced a budget constraint in the experiment.

Following Zizzo's (2002) univariate analysis, we find that subjects played an average of about 16 rounds per dyad. They were tied (the gap was zero) in almost 20% of the rounds, and were either leaders or followers in the other rounds. Zizzo notes that all participants varied their investment efforts, whereas the median investment was 4.5 with a mean of 4.06 and a standard deviation of 2.15. More than 95% of the subjects' decisions were lower than an investment of 7. Participants with an economic or mathematical background had significantly higher investments, while older people invested less than younger people. Even though the experimenters expected participants to run out of money in the later parts of the race, this never occurred; on the contrary, subjects' investments were higher than the median and the mean in the last rounds of the race. More generally, Zizzo found a positive and significant correlation between investment and round.

Some of Zizzo's (2002) results confirm those of Harris and Vickers (1987), but the most important theoretical results do not hold in the experimental study. What was confirmed is a general tendency for investment to increase as the two competitors get closer to each other, but only for the later stages of the race (as the race approaches its end). Zizzo states that this might be the result of an unexplained correlation between investment and progress throughout the race, and not a confirmation of the theoretical paper. Most of Zizzo's results do not confirm the model by Harris and Vickers (1987): leaders did not invest more than followers; the race did not converge to monopoly as the gap between competitors widened but instead followers remained in the race until the gap reached very high levels (5-7 steps); investment did not increase when subjects were close to each other (1-2 steps). Zizzo suggests that the budget constraint present in the experimental study - which is not modeled in the theoretical study - might be responsible for the diverging results and the limited support for Harris and Vickers (1987). I agree with Zizzo’s suggestion, because I believe it makes a large difference whether an agent has a budget constraint or whether its budget is unlimited. Still, Zizzo underlines the fact that real-world R&D departments also face budget constraints and thus the absence of a constraint in Harris and Vickers (1987) is a mere weakness of their theoretical model. Zizzo's main conclusion is that we generally expect a too high degree of rationality on the side of economic agents; instead, participants in the experiment are subject to bounded rationality and cannot make optimal decisions. Zizzo's recommendation for future research is to try and “explain the strong positive correlation between investment and progress in the race”. Zizzo also points out that the contradicting results between theory and experiment “present a puzzle for economists who believe in patent race theory as a suitable approximation to real-world patent races”, a point of view that I am supporting in my dissertation.

Kähkönen (2005) replicates the patent race experiment run by Zizzo (2002), presents its results and compares them to the original experiment. Furthermore, the “data sets of both experiments, the replication and the original, are pooled and analyzed jointly” (Kähkönen, 2005). The heterogeneity between the experiments is considered and Kähkönen asks whether the differences between experiments are just random variation. According to Hunter (2001), replications can be categorized into two main classes: statistical replications and scientific replications. Statistical (or exact) replications have to be identical to the original experiment, in the sense that the exact same sample, variables and procedures are used. Scientific (or close) replications are equivalent to the original experiment, in the sense that the same dependent and independent variables have to be analyzed, the same basic procedures have to be employed, and the sample must come from an equivalent population for the goals and objectives of the study. Kähkönen’s replication is considered to belong to the latter category – it is a scientific replication. Therefore, according to the author, any differences between the experimental results should be due to random variation (sampling variability).

Like in Zizzo (2002), in Kähkönen (2005) a sample of 36 subjects played in dyads a patent race game. The ages and educational backgrounds of the subjects were similar to those in the original experiment. The experimental design followed Zizzo's experiment very closely, with the only exceptions of language (Finnish), currency (Euro), a different experimental software and a different random number generator. The univariate results were strikingly similar to the original experiment: the average duration of the race was 16.6 rounds (versus 16.1 rounds in Zizzo); the percentage of tied rounds was 17.06 % in Kähkönen (versus 18.90 % in Zizzo); investments were slightly higher (but the difference was not statistically relevant, as computed by Kähkönen) in the replication, with median 5, mean 4.51 and standard deviation 1.61, *vis-à-vis* 4.5, 4.06 and 2.15 in the original experiment. Just like in Zizzo, in the replication we still find a positive and significant correlation between investment and round. Even though at first glance it appears that the only difference between the experiments is random variation, a deeper econometric analysis by Kähkönen finds out that the two different data sets cannot easily be pooled and analyzed together, because they show different statistical properties. Kähkönen points out that the unobserved effects in Zizzo were analyzed using a random effects estimator (showing subject-specific effects), while in Kähkönen the random effects model is inappropriate, and a fixed, game-specific effects estimator is used. Eventually, the data were pooled together using a mixed model (featuring both random and fixed effects).

Kähkönen concludes that "the replication was mostly successful and the results of the replication experiment are generally in line with the original one". However, according to the author, the statistically different properties of the two data sets show that the differences between experiments cannot be reduced to just random variation between samples. Kähkönen states that the subject-specific variables were not universally valid across experiments – from the demographic variables only age and educational background were significant in both experiments. With respect to the theoretical paper by Harris and Vickers (1987), Kähkönen, consistent with Zizzo, finds only limited support for the theoretical model. Just as before, we find that the leaders did not always make greater efforts than the followers, the followers did not drop out of the race when the gap increased and the investment efforts did not increase when the gap was very small (in the neck-to-neck or tied situation). Kähkönen recommends that more research and more experiments should be conducted, and more advanced methods to control for the underlying heterogeneity are needed.

Breitmoser, Tan and Zizzo (2010) perform an experimental study on perpetual patent races, based on the original theoretical papers by Aoki (1991), Hörner (2004) and including elements from Harris and Vickers (1987). Perpetual patent races have the characteristic of a "dynamic indefinite" time horizon and they are R&D races featuring "uncertainty and multiple prizes" (Breitmoser et al., 2010). The latter authors emphasize the fact that there are many industries in which firms compete for the development of innovations "without clear finishing lines" (opposite to most patent race models described in Chapters 5.1.1. to 5.1.3., where the finishing line and the exclusive awarding of a patent were an essential feature). In such industries, "innovations are gradual, technology progresses incrementally, and patents are less crucial in defining relative market positions. A leading firm earns more than a lagging firm, e.g. higher product quality justifies a higher price mark-up or captures a larger market share" (idem, p. 446). However, there is no "winner-take-all" feature in this type of competition, and firms coexist in the industry. Relevant examples of such industry structures can be found in the pharmaceutical industry (Cockburn and Henderson, 1994), for semiconductors (Gruber, 1994) as well as for disk drives (Lerner, 1987).

The modelling used by Breitmoser et al. (2010) follows the reasoning used by Hörner (2004), who defines patent races as games consisting of several rounds and with an infinite time limit. In each of the rounds, the participating firms have to choose between high and low R&D effort. As in Harris and Vickers (1987), in Breitmoser et al.'s model high effort is more expensive than low effort, but increases the probability of making a technological advancement. Furthermore, in accordance to the perpetual race model by Aoki (1991), in each round of Breitmoser et al.'s game, the player who is leading in the race earns a larger benefit than the follower.

Based on these features, the model by Breitmoser et al. predicts several types of stationary Markov perfect equilibria. A Markov perfect equilibrium is a set of mixed strategies for all players, which satisfies the following criteria: i) memorylessness (each player's mixed strategy is only dependent on the current state of the game); ii) only the payoff-relevant information is included (and any signals, negotiation or cooperation between players are excluded); iii) the players' strategies form a subgame perfect equilibrium of the game (Maskin and Tirole, 2001). Based on this concept, Breitmoser et al. predict a variety of possible Markov perfect equilibria, which can be characterized into absorbing equilibria, reflecting equilibria and other equilibria. In an absorbing equilibrium, the follower gives up on exerting high effort and thus the position of the leader remains unchanged throughout the race (similar to the concept of ε-preemption which we have encountered in Fudenberg et al., 1983). In a reflecting equilibrium, the leader relaxes and exerts low effort, which allows the follower to catch up by exerting high effort, with the final result that the lead in the race "constantly changes hands" and nobody has an unchallenged position (Breitmoser et al., 2010). Finally, other equilibria are characterized by Breitmoser et al. as being "neither absorbing nor reflecting" equilibria.

Most importantly, Breitmoser et al. underline the fact that Markov perfect equilibria "are highly context-sensitive" and their stability will depend on the parameters of the race, "such as the discount rate, the cost of and probability of progress for high versus low effort, and the benefits from staying ahead" (idem). The main purpose of Breitmoser et al.'s paper is to verify in a controlled laboratory experiment whether the context-sensitivity of the predicted equilibria will hold with real players. Interestingly, they come to a conclusion similar to my results from the experimental Chapter 6: while the theoretical equilibrium predictions are rather unstable, the subjects' real strategies are far less sensitive and follow certain paths. In their paper, Breitmoser et al. employ best response and quantal response models of perpetual patent races, and they compare their predictive performance. Different from a best response model, in a quantal response model: i) subjects have positive probabilities for playing any strategy; ii) "the higher a strategy's payoff, the higher its probability of being played"; iii) subjects "take the "mistakes" resulting from quantal responses into account in equilibrium play" (idem). Breitmoser et al. find that the quantal response equilibrium is a powerful tool to explain the variance in subject responses.

As stated before, according to Breitmoser et al.'s (2010) mathematical framework, a set of two players play a race for an infinite number of rounds. In each round each player can choose between high effort (*H*) and low effort (*L*). The cost of *H* is $c > 0$ and the cost of *L* is normalized to zero. The effort levels *H* and *L* influence their respective probabilities of success α_H and α_L. When a player is ahead in the race, she realizes a payoff *R*; when she is behind, she gets a payoff of $-R$; when the players are tied, each of them has a probability of 50% of being behind or ahead. Players discount future payoffs by a discount rate of δ. By varying α_H and α_L, the authors generate four relevant treatments labelled from A to D. Treatment A is a control treatment in which the high effort strategy is always dominated in all states. All other treatments B, C and D have varying states in which either high effort or low effort should be exerted. The features of Breitmoser et al's experiment are synthesized in Table 5.1.

Table 5.1. Features of Breitmoser et al.'s (2010) experiment

<table>
<tr><td rowspan="2">Players' strategies</td><td>High effort (H)</td><td>Cost $c > 0$</td><td>Probability of success α_H</td></tr>
<tr><td>Low effort (L)</td><td>Cost is zero</td><td>Probability of success α_L</td></tr>
<tr><td rowspan="3">Players' payoffs</td><td colspan="3">R if ahead</td></tr>
<tr><td colspan="3">-R if behind</td></tr>
<tr><td colspan="3">If tied, probability of 50 % of being behind or ahead</td></tr>
<tr><td>Discount rate of future payoffs</td><td colspan="3">δ</td></tr>
<tr><td rowspan="4">Treatments</td><td>$\alpha_H = 0.5$</td><td>$\alpha_L = 0.25$</td><td>Treatment A</td></tr>
<tr><td>$\alpha_H = 0.9$</td><td>$\alpha_L = 0.25$</td><td>Treatment B</td></tr>
<tr><td>$\alpha_H = 0.5$</td><td>$\alpha_L = 0.1$</td><td>Treatment C</td></tr>
<tr><td>$\alpha_H = 0.9$</td><td>$\alpha_L = 0.1$</td><td>Treatment D</td></tr>
</table>

Further in their paper, Breitmoser et al. describe their experimental design. In June 2005 the authors ran a computerized experiment in Frankfurt (Oder), Germany. "Subjects were students from the faculties of Business Administration and Economics, Cultural Science, and Law. A total of 90 subjects participated in the 9 sessions (with 10 subjects per session)" (idem).

Each subject had to play three sessions for each of the conditions B, C and D; each session contained the control treatment A and one of the other three treatments B, C and D. Each session contained ten stages, and subjects were randomly matched at the beginning of each stage. In order to implement the perpetual nature of the race, each stage consisted of several rounds, and in every round the probability that the stage would end was of 10%, which was communicated to all participants (idem).

Descriptive statistics and univariate statistical tests performed by Breitmoser et al. show that the average investment in the experiment was 0.669 and it varied very little between treatments (investment levels were 0.686, 0.611, 0.706 and 0.683 in treatments A, B, C and D). Subjects studying economics "may have invested slightly less", and "there is no evidence of age or gender effects" (idem). There was a significant negative correlation between investment levels and progress in a session. The authors also ran logistic regressions with individual level and session level random effects, taking investment as the dependent variable. Independent variables were: tied situation; leader situation; positive gap; negative gap; stage; stage squared; round; round squared. Results show that "investment tends to be highest at the beginning of each stage, perhaps as players try to grab the lead when faced with new co-players. Increasing the gap between competitors lowers investment in all treatments", and in this case the market converges to a monopoly (idem).

The main results of the experiment run by Breitmoser et al. show that subjects' strategies cannot generally be rationalized. Compared to theoretical predictions, subjects' investment rates were higher than expected, and their "behavior was less context-sensitive" than theory forecasted (idem). "Most of the subjects exert high effort in states where doing so is dominated. […] In all our treatments except the control treatment where low investment was always predicted, competition tended to evolve into an R&D leadership monopoly: a market structure with an entrenched leadership and lower aggregate investment than if competitors stay neck-to-neck" (idem).

Breitmoser et al. point out that it is difficult to assess the validity of equilibrium analysis even in a controlled environment where the optimal strategic behavior can be computed mathematically with great precision.

Breitmoser et al. add to their results the similarities to other empirical studies which found very weak support for the reliability of equilibrium analyses in the case of R&D races (such as Meron and Caves, 1991 or Cockburn and Henderson, 1994). Dealing with a similar problem, Cohen and Levin (1989) note that existing field studies have some fundamental limitations: they cannot measure strategic behavior and they face heterogeneous R&D motivations from the participants in a technology race. Breitmoser et al. suggest that their data is appropriate to explain the failure of the equilibrium concepts in the experiment. One explanation for this failure, provided by the latter authors, could be that the experiment does not allow its participants to gather sufficient learning. Another explanation is that "to players, the long-term payoff implications of different actions may not be sufficiently dissimilar, or saliently so, to justify a sharp distinction between optimal and sub-optimal actions" (idem). Since optimal strategies lack transparency, the concept of quantal response equilibria was employed and provided "a plausible analytical tool" for the experimental data (idem). The authors state that, combined with a rivalry motive, the Markov quantal response equilibrium explains an astounding "95% of the variance in the empirical distribution of responses", while optimal response equilibria are completely unreliable: in most cases, they perform worse than chance (idem). Breitmoser et al.'s recommendations for future research are "allowing for even more learning, varying the number of R&D competitors, reconsidering modeling assumptions, and analyzing welfare and policy implications" (idem).

The main design features of the experiments I have discussed in this chapter are summarized in Table 5.2.

Table 5.2. Main design features of the experiments from Chapter 5.2.1.

Silipo (2005) • tested the theoretical paper by Fudenberg et al. (1983) • model of a multi-stage race with costly efforts to accumulate units of knowledge over time • first player to achieve $N = 30$ units of knowledge wins the race and is awarded the patent • number of subjects: 172, who played 86 races in dyads • effort levels: 0, 1 or 2 units of knowledge with convex costs \$ 0,

\$ 0.44 or \$ 1.11 • each round consisted of two stages: stage 1) decision on whether to cooperate or invest alone; stage 2) investment decision • if both players cooperate throughout the race and reach 30 steps jointly, they split the prize minus the total costs • between-subjects design with six treatments generated by varying the prize (high or low) and the initial positions of the two players: (0,0); (0,1); (0,2) • the findings of the paper contradict the original paper by Fudenberg et al. (1983)
Zizzo (2002) • tested the paper by Harris and Vickers (1987) • participants were 36 subjects in 9 sessions à 4 subjects who played in dyads a “prize race” • in every race, players were endowed with 500 points, the prize was set to 1000 points, and the decisions on investment for player 1 and 2 were x and y, with probabilities of winning the round $x/x+y$ and $y/x+y$ • convex cost of investment, with $c(x) = x^2$, which was subtracted from each subject’s endowment • after the decision, information on the winning probabilities, investment of each player and the winner was provided • the winner made one progress step in each round • the first player to complete 10 progress steps became the winner of the race and received the prize of 1000 points, from which the investment costs were subtracted • results showed only limited support for the theoretical predictions by Harris and Vickers (1987) and most of the predictions did not hold in the experimental study
Kähkönen (2005) • replication of the patent race experiment by Zizzo (2002) • a sample of 36 subjects played in dyads a patent race game; the subjects had ages and educational backgrounds similar to those in the original experiment • the experimental design followed Zizzo’s experiment very closely, with very few exceptions • results were strikingly similar to those in the original experiment; however, the differences between the two experiments cannot be

reduced to random variation between samples, but to the different properties of the two data sets
Breitmoser et al. (2010) • tested the models by Aoki (1991), Hörner (2004) and Harris and Vickers (1987) • a set of two players play a race for an indeterminate number of rounds • in each round, each player chooses between high effort *H* and low effort *L*; the effort levels *H* and *L* influence the probabilities of success in a round α_H and α_L • when a player is ahead in the race, she realizes a payoff *R*; when she is behind, she gets a payoff of -*R* • treatments or conditions A, B, C and D were generated by varying the probabilities of success α_H and α_L • 90 subjects participated in 9 sessions à 10 subjects • each subject played 3 sessions for each condition B, C and D, while condition A was included in each session as control treatment • each session contained ten stages, and subjects were randomly re-matched at the beginning of each stage • each stage consisted of an indeterminate number of rounds, and participants were informed that in each round the probability that the stage would end was of 10 % • the main results of the experiment show that subjects' strategies cannot generally be rationalized; subjects' behavior was less context-sensitive than theory forecasted

To sum up my conclusions from Chapter 5.2.1, I found that the predictions of the theoretical patent race literature are inconclusive when the models are tested experimentally. Indeed, only four experimental tests of traditional patent races have been performed, and I have described them extensively in this chapter. First, I found that the theoretical model of Fudenberg et al. (1983) is not supported by the experimental study of Silipo (2005). Here the problem was that Silipo significantly modified Fudenberg et al.'s original research question by introducing the cooperative option, and therefore it is no wonder that the results of the two studies contradict each other. Second, there are the contradictory results between the prestigious theoretical model by Harris and Vickers (1987) and its experimental tests in Zizzo (2002) and Kähkönen (2005). Here, the experiments were designed extremely close to the theory (with the exception of a budget constraint, which was needed in the experiment and represented a reasonable assumption). Yet, Harris and Vickers' predictions about investment behavior as a function of the gap between competitors generally do not hold when tested experimentally. As Zizzo points out, these contradicting results severely undermine the predictive power of the patent race literature as a "suitable approximation to real-world patent races". Third, the theoretical work on perpetual races elaborated by Aoki (1991) and Hörner (2004) does not produce the expected predictions in the experiment run by Breitmoser et al. (2010). Here, theoretical models are very sensitive to parameter changes, while participants' real investment behavior in an experimental race is not context-sensitive at all, and the authors conclude that the subjects' decisions are difficult to rationalize.

For the sake of completion, the next chapter presents experiments with patent races not rooted in the traditional literature (experiments based on own theoretical models), and it adds to an already very high level of complexity.

5.2.2. Experiments Testing Their Own Theoretical Models

In this chapter, I will chronologically present four studies - published in highly ranking scientific journals - which develop own theoretical models of R&D or patent racing and then test these models experimentally. I leave out unpublished papers dealing with patent race experiments (e.g. Cantner et al., 2004 and 2005, among others). Towards the end of the chapter I will formulate my general conclusions from Chapter 5.

The study by Isaac and Reynolds (1992) performs a series of laboratory experiments in the context of R&D racing. In the experiments, a small number of sellers have to compete by making decisions about pricing, production and cost-reducing R&D. The sellers receive rewards for their innovative behavior, but these rewards depend strongly on the type of competitive market: large differences can be observed in a monopoly market versus a four-seller competitive market. The type of competitive market also has a significant effect on output and pricing of the innovation: "aggregate R&D is higher under competition than under monopoly and prices follow marginal cost reductions much more quickly under competition than under monopoly", strongly resembling a type of Schumpeterian competition (idem). The experimental results are in high accordance with theoretical predictions.

Isaac and Reynolds' (1992) study is mainly based on the concepts of market evolution and non-price competition, which were introduced by Schumpeter (1950). Schumpeter's view of the competitive process is that of a dynamic process: under firms' innovative efforts, markets continuously change and evolve; not price competition matters, but the firms' underlying competition in technology, R&D and quality. This type of competition "strikes not at the margins of the profits and the outputs of existing firms but at their foundations and their very lives" (Schumpeter, 1950, p. 84). Furthermore, Isaac and Reynolds build up their theoretical approach based on important studies about the relationship between R&D and market structure, on a careful selection of studies dealing with competition and cost reduction, and on previous research on the topic of stochastic innovation (which has a branch in the patent race literature through the papers by Loury, 1979 and Lee and Wilde, 1980, but was further extended through the studies by Mortensen, 1982 and Stewart, 1983). According to stochastic innovation theory, each of *N* firms simultaneously chooses a level of R&D investment, and their choices "generate probabilities of innovative success for each player" (Isaac and Reynolds, 1992) resulting in distinct rewards for successful and unsuccessful players (similarly to Harris and Vickers, 1987). The levels of rewards are affected by factors such as patent protection, imitation and the competitive nature of the market (Isaac and Reynolds, 1992). The study by the latter authors goes beyond the stylized facts of the stochastic innovation theory. Isaac and Reynolds develop a framework in which sellers make decisions on pricing and output in a product market. These decisions, together with the product buyers' decisions (which were computer simulated), endogenously determine the sellers' private benefits to cost-improving R&D. The authors write that "the experimental environment is explicitly dynamic. Price and output decisions are made over a sequence of market periods and cost reductions for a seller are cumulative over the course of an experiment" (idem). Isaac and Reynolds point out that this dynamic framework allows them to study the development of price, marginal cost and market share during an experiment, based on subjects' behavior.

In the following, I describe the implementation of the experiments by synthesizing information from Isaac and Reynolds' paper. In the experiments, sellers played a number of fifteen periods. During the first five periods, sellers had no opportunity to perform R&D in order to reduce their costs (the purpose of this was "to familiarize the subjects with the cost and demand conditions", idem). Starting with the sixth period, sellers participated in an innovation stage followed by an output market stage. In the innovation stage, sellers could purchase draws (with a cost of 10 cents per draw) from a bingo cage with balls labeled from 1 to 10. If the ball with the number 10 was among the player's draws, the seller was considered an innovator and her marginal costs were reduced by 25 cents per unit. The levels of R&D investments and their success were made public before the subsequent (market) stage began. Isaac and Reynolds (1992) ran twelve experiments: two experiments were controls, each consisting of four sellers without the possibility of innovation; six experiments had four-seller markets with the possibility to innovate; the remaining four experiments were monopoly control experiments (in which there is only one seller in the market, but she can still choose to innovate).

The results of the experiments by Isaac and Reynolds confirm the theoretical strength of the concept of Schumpeterian competition. When they find themselves on competitive (four-seller) markets, sellers engage in costly innovation and when they succeed, prices start falling and the industry gets more concentrated. To make sure that these results indeed represent Schumpeterian competition, the authors compare the results of competitive market experiments with those on monopoly markets. The data strongly indicates that in all monopoly settings, "final market prices are above the original competitive equilibrium prices" (idem). Furthermore, the competitive markets generated significantly higher levels of total welfare, and almost all this welfare is transferred to consumers; on monopoly markets, total surplus is lower, and most of the surplus accrues to the monopolist. Average aggregate investment was also found to be much greater in the four-seller markets *vis-à-vis* the monopoly markets (due to the fact that the monopolist is not under pressure to innovate, while the competitive sellers lose market share if they do not). All in all, Isaac and Reynolds's (1992) study is based on a solid theoretical foundation (that of Schumpeterian competition), and the results of their experiments show strong support for the theory.

Sbriglia and Hey (1994) performed an exploratory study in the field of R&D competition through a series of experiments. The purpose of their study was to find out "how firms search for innovations that will allow them to gain significant profits, as in the case of patentable innovations, and how they adjust their search strategies in order to take account of the competition from other potential innovators" (idem). Further aspects under investigation mentioned by the authors were related to the distribution of firms' spending during the search process, the role of information in the experiments and factors that would influence firms' decisions to enter or exit the competition. The experiments were designed as innovation races in which N subjects competed "to discover an unknown combination of elements" that represented a "patentable innovation" (idem). A subject who made the discovery alone was awarded the whole prize V, or she had to share the prize equally with other players if they made the discovery simultaneously. Subjects who did not make the discovery had to pay for their own search costs (which were subtracted from a participation fee). The unknown combination of elements was represented by n letters of the alphabet (e.g. ZDNFVKL with $n = 7$), which were randomly selected from a set of m letters. Sbriglia and Hey suggest that this design is reflective (albeit in a highly simplified form) of the research process in the pharmaceutical industry. In order to find the correct combination, subjects could buy "trial combinations of n letters" in each round and were informed about which letters were correct, at the end of each round (idem). As stated by the authors, 57 subjects participated in 15 total experiments, including 3 pilot studies, 7 experiments with a deterministic information structure and 5 with stochastic information. The information structure was linked to whether a player had knowledge about other players' search performance perfectly or stochastically.

The main results from Sbriglia and Hey's (1984) study are related to the patterns of behavior for different players, who were categorized into winners and losers. The authors write that "in all experiments, the winners applied a successful search procedure and did not play randomly". Generally, winners "invested more than their participation fee" and also more than the losers (idem). The authors found that losers played less aggressively or were less lucky with their search process. The experiment also allowed subjects to decide not to enter the competition at all. A questionnaire asked players about their reasons not to enter, and found out that some subjects decided not to enter because of “lack of a specific search strategy” and because they preferred the certain participation fee to the “uncertain investment necessary to find the unknown combination” (idem). I believe that non-entering subjects acted so because of their risk attitudes (that is, risk aversion), but one cannot test this hypothesis because Sbriglia and Hey did not measure risk attitudes in their experiment. Finally, the role of deterministic versus stochastic information was analyzed and the authors concluded that subjects largely ignored their opponents’ search performance and were mainly concerned with their own speed in the race. Sbriglia and Hey generally conclude that their experiment offers a realistic framework of research competition for a patentable innovation, and that it provides useful insights into the link between strategy, skill, information and incentives to enter or quit the race as well as timing and level of investment.

The important contribution by Amaldoss and Jain (2002) was concerned with an experimental investigation of an asymmetric mixed-strategy game in the context of a patent race. The authors start their paper by emphasizing the importance of the mixed-strategy Nash equilibrium concept, which has been used, among others, in R&D and pricing models. Its popularity is, as stressed by the authors, due to the fact that a mixed-strategy equilibrium often exists when there are no pure-strategy equilibria in a game. Amaldoss and Jain point out to empirical evidence which reinforces the concept: in laboratory experiments, the mixed-strategy equilibrium was shown to account well for subjects’ behavior in symmetric settings but it leads to theoretical results that are counter-intuitive in asymmetric settings. Amaldoss and Jain’s study contains an asymmetric setting of two firms which race “to develop a product and obtain a patent”. The asymmetry lies in the fact that one of the two firms “values the patent more because of its market advantages, such as brand reputation and distribution network” (idem). The authors write: “Intuition would suggest that the firm that values the patent more will invest more and, consequently, win the patent more often. However, our theoretical results show that, in fact, in equilibrium the opposite is true: The player who has less to benefit from winning the patent will invest more aggressively and win the patent more often” (idem). Mathematically, the reason for these surprising results is intimately connected to the nature of a mixed strategy: the equilibrium strategy of a player i is computed based on the payoffs of the opposing player, player j, and the other way around. As one can also find in Chapter 6.1.3, this happens because player *i* randomizes between its strategies so as to make player *j* indifferent between all of player *i*'s strategies, and vice versa. Amaldoss and Jain acknowledge that in their study this mathematical result also has a plausible real-world interpretation: a market advantage on the side of the market leader "can provoke aggressive investment by competitors and potentially impede innovation" by the leader (idem). Therefore, according to the authors, the failure to innovate by the leader is not due to management inefficiencies (or naïve, nonstrategic players), but it arises from strategic behavior (or perfectly rational players).

In the following, I briefly describe verbally and mathematically the model by Amaldoss and Jain (2002). Here, two firms on the market competitively invest in R&D in order to win a patent. Any firm can invest an amount $x \in [0, c]$, where c "represents the financial constraints imposed on the firm by the capital market" (idem). The investment is made in increments of c, so a firm can invest $(0, c/k, 2c/k, 3c/k, \ldots, c)$ units (idem). Like in Gilbert and Newberry (1982) as well as Fudenberg et al. (1983), "the firm that invests more wins the patent" (Amaldoss and Jain, 2002). Firms are asymmetric in the valuation of the patent: one firm (firm *L*) has a lower payoff from winning the race, denoted as r_L, than the other firm (firm *H*), which writes a payoff of r_H, and $r_H > r_L$. Furthermore, $c < r_L$. The financial constraint c and the two firms' payoffs r_i are common knowledge. If both firms invest the same R&D amount x_i, Amaldoss and Jain make the assumption that the "firms engage in Bertrand competition and both firms make zero profits from the patent" (idem). The authors model firm *i*'s payoff as:

$$\Pi_i(x_i) = \begin{cases} r_i - x_i & \text{if } x_i > x_j;\ i \neq j, \\ -x_i & \text{otherwise.} \end{cases} \qquad \textbf{(5.1)}$$

I found it quite a critical assumption that the payoff of any firm was modeled as $-x_i$, if firms invest the same amount x_i. I cannot find any link to real-life situations in which firms would suddenly make a loss if they invested the same amount of R&D expenditures as another firm. However, this working assumption is essential to Amaldoss and Jain's paper. In the following, the authors characterize the equilibrium solution. They show in their Proposition 1 that the unique equilibrium of the game is "for Firm *H* to invest *ic/k* discrete units of capital (where $i = 0, 1, \ldots, k$)" with the following probability:

$$p_H\left(i\frac{c}{k}\right) = \begin{cases} \frac{c}{k}\frac{1}{r_L} & \text{if } i = 0, 1, \ldots, k-1, \\ 1 - c/r_L & \text{if } i = k. \end{cases} \tag{5.2}$$

Similarly, for Firm L the probability is:

$$p_L\left(i\frac{c}{k}\right) = \begin{cases} \frac{c}{k}\frac{1}{r_H} & \text{if } i = 0, 1, \ldots, k-1, \\ 1 - c/r_H & \text{if } i = k. \end{cases} \tag{5.3}$$

Amaldoss and Jain show formally how Proposition 1 leads to the following results: 1) "Firm L's equilibrium strategy does not depend on r_L, but rather depends on r_H"; 2) "on average, Firm L invests more than Firm H"; 3) "Firm L is more likely to win the patent" (idem). The authors point out again that these three results "are not consistent with intuition, but are the only results consistent with strategic thinking about the game" (idem). No other propositions or results were formulated.

In the next section of their paper, the authors proceed with the experimental test of the theoretical model. Their main research question is whether real subjects will play according to normative equilibrium predictions, or will apply a heuristic with no normative basis (of the type: "high-reward subjects should invest more than low-reward subjects", idem). Participants to the experiment were 36 student subjects. They played in two groups of 18 subjects each, generating two versions of the same experiment. In each experiment, the 18 subjects were "randomly divided into nine low-reward players and nine high-reward players, and their reward condition remained the same throughout the experiment" (idem). Each subject played 80 trials of the same game, all subjects were randomly re-matched at the beginning of each trial (randomly changing opponents) and each subject was given feedback on both players' investments, the winning player and both payoffs at the end of each trial. The parameters of the experiment are summarized in Table 5.3.

Table 5.3. Parameters of Amaldoss and Jain's (2002) experiment

Financial constraint	$c = 2$
Basic increment	$k = 2$
Reward for the high-reward player	$r_H = 7$
Reward for the low-reward player	$r_L = 4$
Endowment at the beginning of each trial	2 francs
Possible investments	0, 1 or 2 francs
If subjects invested the same amount x_i, each received a payoff of $-x_i$	

The results of the study by Amaldoss and Jain showed that, on the aggregate level, both the low-reward subjects and the high-reward subjects conformed to the predictions of the theoretical model, which shows the power of the mixed-strategy equilibrium concept. There were some deviations from the normative predictions, but only for a subset of the high-reward subjects, and only for the first 40 trials; in the last 40 trials the behavior of these subjects converged to the predictions, which I interpret as learning effects. As expected, low-reward subjects played more aggressively and invested more than high-reward subjects. However, although aggregate results show support for the theory, individual decisions do not: one still finds that low-reward subjects invested more, but there is considerable heterogeneity in all subjects' decisions and the distribution of investments significantly differed from the model predictions. In order to explain individual decisions, Amaldoss and Jain employed an adaptive learning model, namely the experience-weighted attraction learning model by Camerer and Ho (1999). Reasons given by Amaldoss and Jain for the selection of the learning model are that it allows for learning effects, and that it was applied to various situations, showing good predictive power. The way the experience-weighted attraction learning model works is the following: i) the past success of a strategy will make it more likely for the strategy to be chosen again; ii) "unchosen strategies, which might have yielded high payoffs, are more likely to be chosen in the future"; iii) "with experience, players move to reduce discrepancies between actual and foregone payoffs"; iv) there is a forgetting parameter in the sense that previous decisions are discarded after some time (idem). After calibrating and applying Camerer and Ho's learning model to the individual decisions, Amaldoss and Jain found a very good model fit: indeed, subjects learn and adapt throughout the experiment.

To sum up all results from Amaldoss and Jain's paper, one finds that the research tested "the descriptive validity of some of the counterintuitive implications of using mixed strategy when players are asymmetric" (idem). The authors developed a patent race model, tested it experimentally and received good support for the theory on the aggregate level. The heterogeneity of decisions on the individual level could be well accounted for by employing an adaptive learning model. The model provides an interesting explanation of the reasons why a firm with a market advantage might fail to innovate: the advantage of one firm provokes strategic aggressive investments by its competitors and might impede the firm's innovative ability. Amaldoss and Jain point out that this result is also consistent with some field observations: Lerner (1997) analyzed innovative behavior in the disk-drive market during the period between 1971-1995 and found out that market leaders innovated at a lower rate than their competitors, and they often lost their leading position to market challengers. All in all, the paper by Amaldoss and Jain (2002) has a very strong theoretical basis, a good experimental design as well as excellent results and implications, which are also supported by empirical observations.

Nieken and Sliwka's (2010) paper studies risk taking in a two-person tournament by creating a theoretical model and performing an experiment. Tournaments are a general case of patent races since contestants compete "for a limited set of prizes" and the best performant "receives the winner prize", while "less successful competitors only receive lower loser prizes" (idem). Similarly to the experiment described in Chapter 6, in Nieken and Sliwka's paper two agents simultaneously choose between a risky and a safe strategy. As the authors put it, this sort of situation has relevance in the real world, most notably for "mutual fund managers" who "have to decide between risky and safe portfolios", firms which have to choose between the introduction of a new technology and staying with the *status quo* or also politicians who "gamble for resurrection" by choosing risky policy alternatives for fear that they might lose the election (idem). The main departure of Nieken and Sliwka from other models is that they allow contestants to imitate each other's risky strategies: for example, if a competitor who has remained behind in a race follows a risky strategy, "the front runner may well try to adopt a similar strategy" in order to protect her lead (idem). In other words, Nieken and Sliwka "allow for different degrees of correlation between the outcomes of the risky strategies of the contestants", and their results strongly depend on the level of this correlation (idem). The two agents involved in the tournament have asymmetric positions (like in Chapter 6), so one of the contestants has a lead over the other. The winner of the tournament receives a prize with a normalized value of one, while the loser's value is set to be zero. Similarly to the theoretical model I present in Chapter 6, Nieken and Sliwka's model also computes three Nash equilibria: i) the leader chooses the safe strategy and the follower the risky strategy; ii) they both choose the risky strategy; iii) they both play a mixed-strategy equilibrium. The difference to this dissertation is that in Nieken and Sliwka, a stronger correlation between the risky strategies means that it becomes more attractive for the leader of the race to choose the risky strategy. In the following, I provide the mathematical description of Nieken and Sliwka's theoretical model.

The authors consider a tournament between two agents *A* and *B* who simultaneously decide among a risky and a safe strategy, and their decision is depicted as $d_i \in \{r, s\}$ for $i = A, B$. Each agent's decision influences the distribution of her performance, denoted as y_i, as follows:

$$y_i = \mu_s \qquad \text{if } d_i = s \tag{5.4}$$

$$y_i = \tilde{y}_i \sim N(\mu_r, \sigma^2) \quad \text{if } d_i = r \tag{5.5}$$

It is furthermore assumed that "the performance indicators $\tilde{y}_i$ are correlated with a correlation coefficient $0 \leq \rho \leq 1$" (idem). Such a coefficient does not exist in my experimental model from Chapter 6, because in my model the players' strategies are completely unobserved and simultaneous. Turning back to Nieken and Sliwka's model, Agent A has a lead over agent B, and this lead is denoted as $\Delta y_A \geq 0$. The authors introduce $\Delta\mu = \mu_r - \mu_s$, and $\Delta\mu$ "is positive if the risky strategy has a higher expected outcome than the safe one and negative in the opposite case" (idem). The authors then show that *(risky, safe)* and *(safe, safe)* can never be Nash equilibria and proceed with the mathematical characterization of the three existing equilibria through their Propositions 1-3. Proposition 1 states that "a pure strategy Nash equilibrium exists in which the leading player A plays the safe strategy and player B plays the risky strategy if and only if":

$$\frac{\Delta\mu}{\Delta y_A} \leqslant 1 - \frac{1}{\sqrt{2(1-\rho)}} \tag{5.6}$$

According to Proposition 2, "a pure strategy Nash equilibrium exists in which both players choose the risky strategy if and only if":

$$\frac{\Delta\mu}{\Delta y_A} \geqslant \max\left\{1 - \frac{1}{\sqrt{2(1-\rho)}}, \frac{1}{\sqrt{2(1-\rho)}} - 1\right\} \tag{5.7}$$

Finally, Proposition 3 states that "a Nash equilibrium in mixed strategies exists if and only if":

$$1 - \frac{1}{\sqrt{2(1-\rho)}} < \frac{\Delta\mu}{\Delta y_A} < \frac{1}{\sqrt{2(1-\rho)}} - 1 \tag{5.8}$$

Furthermore, Nieken and Sliwka point out in their Proposition 3 that "in any mixed strategy equilibrium, player A chooses the risky strategy with a higher probability than player B if the risky strategy leads to a higher expected outcome than the safe one. If $\Delta\mu < 0$ player B chooses the risky strategy with a higher probability than his opponent".

The experimental design of Nieken and Sliwka (2010) contained three different treatments in which the correlation coefficient ρ of the risky strategy was varied and assumed values of zero, one and 1/2. For each treatment one session with 24 participants was implemented. Each participant played five trial periods plus 23 payoff-relevant periods and was randomly matched with another partner in every period. In any period, every subject was endowed with a score that was "drawn from a normal distribution with a mean of 150 points and a standard deviation of 42 points" (idem). This endowment generated a leader, a follower, and a gap between them. Any subject had to choose between a safe and a risky strategy. The safe strategy guaranteed the player 80 additional points, while the risky strategy generated additional points that were drawn "from a normal distribution with a mean of 100 points and a standard deviation of 20 points" (idem). The final number of points determined the winner and the loser. Feedback was provided at the end of every period. After the experiment was completed, one of the 23 periods was randomly selected and entitled the winner with 25 Euros, while the loser earned only 5 Euros. An additional show-up fee of 2.50 Euros was paid to all subjects.

Nieken and Sliwka's (2010) results were pretty clearly cut: the possibility that contestants imitate the risky strategies of their opponents (and that the outcomes of the policy choices are thus correlated) strongly influenced the equilibrium of the game. The key predictions of the model were confirmed: if the outcomes are uncorrelated, leaders choose the safe strategy more often than the followers, but the opposite is valid if the outcomes are perfectly correlated. The authors point out that their results are relevant for "gambling in competitive situations. When the competitors have access to similar policies, the correlation between the outcomes of the risky strategies will be high" and therefore leaders will have higher incentives to choose the risky strategy (idem).

The most important features of the experiments from the current subchapter are synthesized in Table 5.4.

Table 5.4. Main design features of the experiments from Chapter 5.2.2.

Isaac and Reynolds (1992) • tested their own developed model of Schumpeterian competition • in their experiment, sellers compete by making decisions on pricing, output and cost-reducing R&D • the authors ran 12 experiments: 2 experiments of four-seller markets without the possibility to innovate, 6 experiments of four-seller markets with the possibility of innovation, while 4 experiments were monopoly control experiments with the possibility to innovate • in the experiments, sellers played 15 periods, in which the first 5 periods had no R&D reduction possibility, and starting with period 6 sellers had an innovation stage and an output market stage • in the innovation stage, sellers could purchase draws that could lower their marginal costs; in the market stage, buyers' decisions were

computer simulated • the results of the experiment confirmed Schumpeterian competition, in that welfare is higher and prices lower on competitive markets compared to monopoly markets
Sbriglia and Hey (1994) • exploratory study in the field of R&D competition • the experimental design was of innovation races in which *N* subjects competed to discover an unknown combination of randomly chosen alphabet letters • a subject who made a discovery alone was awarded a prize *V*, or had to share the prize with other discoverers • 57 subjects participated in 15 total experiments which included 3 pilot studies, 7 experiments with deterministic information and 5 experiments with stochastic information • results showed that winners applied successful search strategies, invested more than their participation fee and more than losers • generally, the experiment provided insights into search strategies for patentable innovations
Amaldoss and Jain (2002) • experimentally tested an asymmetric setting of two firms which race to develop a product and obtain a patent • the authors tested a mixed-strategy equilibrium analysis in which the player with a lower patent valuation is expected to act more aggressively that the leading player of the race • participants were 36 students subjects, who played in two groups of 18 subjects each, generating two replications of the same experiment • in each experiment, the 18 subjects were divided into 9 low-reward and 9 high-reward subjects who remained unchanged in their roles throughout the experiment • any player could invest increments of a financial constraint *c*, and the player investing more won the patent • each subject played 80 trials of the same game with random re-matching in each trial and feedback on investment levels, winning player and payoffs after each trial • on the individual level, an experience-weighted attraction learning model could explain subjects' decisions • on the aggregate level, both the low-reward and the high-reward

subjects conformed to the predictions of the theoretical model, which emphasized the power of the mixed-strategy equilibrium concept
Nieken and Sliwka (2010) • in their experimental setting, two agents with asymmetric positions (in the sense that one agent has a lead over the other in a tournament) simultaneously decide between a risky and a safe strategy • the winner of the tournament received a prize with a normalized value of one, while the loser's utility is set to zero • the experimental design contained three different treatments in which the correlation coefficient of the risky strategies took values of 0, 1 and 1/2 • for each treatment, one session of 24 participants was conducted, and each participant played 5 trial periods plus 23 payoff-relevant periods with random re-matching in each period • in each period, subjects were endowed with a value drawn from a normal distribution; the safe strategy guaranteed the player a certain number of points, while the risky strategy was drawn from a normal distribution; the final number of points determined the winner and the loser, and feedback was given at the end of each period • the experimental results confirmed the key predictions of the model: if outcomes are uncorrelated, leaders choose the safe strategy more often than followers, but the opposite is true if the outcomes are perfectly correlated

A good theory shows its power when it is confronted with experimental testing and when, time and time again, the theory produces reliable and consistent predictions. As one can see throughout Chapter 5, the opposite is true about the patent race literature (and thus I would like to underline that it will probably always remain "the patent race literature" and not a patent race theory). There are wild contradictions and extreme complexity within the theoretical literature (Chapters 5.1.1 to 5.1.5). The contrast between theory and experiments (Chapter 5.2.1) proves to be even starker and wipes out the results of the literature. The further experiments on patent races (Chapter 5.2.2) only increase the complexity and do not lead to any resolutions. However, a few excellent experimental studies can be found in this last subchapter.

In this context, I would like to build on the results by Breitmoser et al. (2010), who pointed out that models are highly sensitive to parameter changes, while people are less context-sensitive. I will design a simple and intuitive game-theoretical model which includes important features of the patent race literature, and which also allows for large ranges of parameters under which theoretical predictions should be stable. I will then perform an experimental test of the model and verify the consistency of real-player decisions with the theory. This endeavor is the topic of Chapter 6.

CHAPTER 6

Patent Races in the Experimental Laboratory

"One cannot, without empirical evidence, deduce what understandings can be perceived in a nonzero-sum game of maneuver any more than one can prove, by purely formal deduction, that a particular joke is bound to be funny."

Schelling (1960)

"Experimental economics seeks to establish a general theory linking economic factors, such as incentives, rules, and norms, to behavior."

Camerer (2011)

"In entrepreneurship many objects and relationships to be researched are dynamic or are embedded in a dynamic environment. These dynamics potentially threaten the reliability of ostensibly identified relationships in field studies. Only with experimental control, the factors of interest may be discriminated from "noise" of other rapidly changing factors."

Schade and Burmeister-Lamp (2009)

The previous chapter showed how a large body of literature on patent races has so far yielded conflicting and unpredictable results. Small changes in model assumptions led to chaotic outcomes, experiments showed no support for theoretical predictions and theoretical papers either contradicted each other or analyzed incompatible research questions. However, there were exceptions to this inconclusive literature, in the form of the experiments based on own theoretical models, which I have described in Chapter 5.2.2. Most notably, the papers by Amaldoss and Jain (2002) as well as Nieken and Sliwka (2010) asked essential questions about the nature of two-player patent races, designed models based on established theoretical constructs and answered their research questions with the help of experimental tests. My experimental study from this chapter is closest to the research these authors performed, and sheds new light on the delicate subject of patent races.

Therefore, I take a novel research approach by designing and experimentally testing a more intuitive game-theoretical model of a patent race. The starting point for my endeavor is the prestigious publication by Nalebuff (1988), in which a patent race is modelled similarly to a sailing competition[7]. In Chapter 6.1, I show my contribution to the existing literature and I adapt the example of a sailing race to match the realities of a model of R&D competition. I use a numerical equilibrium analysis to arrive to an appropriate parameterization of a patent race. Chapter 6.2 starts by showing the importance of experiments for economic research, then formulates propositions and hypotheses and finally presents the experimental design and implementation. Chapters 6.3 and 6.4 provide the econometric analysis and the results of the experiment.

6.1. From a Sailing Race Puzzle to a Game-Theoretic Patent Race Model

6.1.1. Nalebuff's (1988) Sailing Race

Nalebuff (1988) describes the paradox situation encountered during the 1983 America's Cup finals in sailboat racing. The American skipper Dennis

---

[7] I am thankful to Avichai Snir for his suggestion to use Nalebuff's (1988) model.

Conner on board the *Liberty* boat was facing the Australian skipper John Bertrand who was in the lead of *Australia II*. The clear favorite of the race was the American: after the first four races of the Cup, Dennis Conner's *Liberty* was leading “3-1 in a best out of seven series” (idem). On the morning of the fifth race, as reported by The New York Times (September 22, 1983), “cases of champagne had been delivered to *Liberty*'s dock. And on their spectator yacht, the wives of the crew were wearing red-white-and-blue shorts, in anticipation of having their picture taken after their husbands had prolonged the United States' winning streak to 132 years.”

But, as Nalebuff describes, “it was not to be”. *Liberty* started with a leader advantage of 37 seconds when *Australia II* “jumped the gun and had to recross the starting line” (idem). Being at disadvantage from the start, the Australian skipper decided to take a risky strategy: it sailed “far to the left of the course in the hope of catching a wind shift”, while the American boat kept to the safer route on the right hand side of the sailing course (idem). As Nalebuff puts it, the “gamble paid off. The wind shifted five degrees in *Australia II*'s favor and they won the race by one minute and forty-seven seconds. Two races later, *Australia II* won the series”.

Nalebuff (1988) then suggests the main guidelines according to which one could apply the same principles in designing a patent race model. Two firms could be entrenched in a competition for the development of an R&D project. Only the first firm to complete the project is awarded a patent (which should ensure rich profits by capturing the market). One firm is six months ahead of its competitor. There are two strategies any firm can pursue, RISKY or SAFE. The strategy SAFE takes two years but is guaranteed to bring the developer to a successful completion. The strategy RISKY takes only one year, but there is a 50% chance that it will work, and a 50% chance that the R&D project will fail. According to Nalebuff's design, a firm that fails with the RISKY strategy has to return to the SAFE strategy and take an additional two years to complete. Each firm has the same two strategies to choose from: RISKY or SAFE. Any player's chance of success is independent of whether the competitor's RISKY strategy works or fails.

Nalebuff explicitly states that there is no time discounting in this model. I found this assumption critical and I will modify it in the adapted patent race model. Nalebuff also underlines one significant departure from the sailboat race, in that the two sailboats can observe each other, whereas firms in a patent race cannot. Indeed, in an R&D race each firm's move is independent and unobservable. Another important assumption made by Nalebuff, which I found reasonable and adopted in my later model, is that due to limited resources a firm cannot pursue both strategies simultaneously.

Nalebuff asks several research questions about this patent race. Which strategy maximizes a player's chance of winning? What will be the counterpart's behavior? How important is secrecy about a player's own decision? In the following, I will answer these questions as well as add new questions such as: how large should the prize be, whether one should include cost considerations and time discounting factors, or what happens if one varies the probability of success for the RISKY strategy.

Overall, I found Nalebuff's puzzle to be very appropriate for the purpose of designing an intuitive game-theoretical model that I can then test experimentally. The puzzle contains many of the features that were quite intensely scrutinized in the "traditional" patent race literature:

- it is a game-theoretical model, which was endorsed by Fudenberg et al. (1983) and by Reinganum's (1984) article;
- it is a "tough" patent race: there is a single prize, a "winner-take-all" setting with one winner and one loser, as in Reinganum (1982), in the model on ε-preemption by Fudenberg (1983), as well as in Harris and Vickers (1987), Grossman and Shapiro (1987) and Park (1987). The toughness of the model implies that it is particularly suited to bring forward the competitive behavior of individuals;
- it is an asymmetric game, which was the concern of a large stream of the patent race literature. Reinganum (1989) identifies four streams of this literature, which are described in Chapter 5.1.3. Out of these four streams, asymmetric races stand out as especially relevant for markets with an incumbent monopolist and a challenger, but have further implications for our general understanding of competitive behavior. Furthermore, asymmetric races represent the only stream of literature (out of the four streams identified by Reinganum, 1989) which has been tested experimentally. Reinganum (1989) showed that, in a deterministic innovation setting, the leader invests more than the follower, because the first actor knows she can keep her

market position for a long time, or forever. Differently, if innovation is stochastic, the role of the leader circulates around the industry, and thus followers have an incentive to invest and challenge the position of the leader. My experiment, based on Nalebuff's puzzle, includes both a deterministic and a stochastic innovation setting, and therefore aims at core questions of the asymmetric patent race literature;
- Cockburn and Henderson (1994) found it "a daunting prospect" to match the results of the theoretical patent race literature with empirical evidence. Since empirical evidence is nearly impossible to find for the standard patent race models, the only way to test the features and assumptions of such models is through experiments.

The relevance of Nalebuff's puzzle becomes even more obvious after I finish building the theoretical model at the end of Chapter 6.1. Even though I could not anticipate it *ex ante* based on the original puzzle, the final model reflects important research questions that come up in prestigious papers. In the following, I show my contribution to the existing literature:
- the issue of experimentally testing asymmetric mixed-strategy equilibria in the context of a patent race, which I am also doing in this chapter, was performed in the experiment by Amaldoss and Jain (2002), which I have described in Chapter 5.2.2. However, my main departure from these authors' paper is that they only model a mixed-strategy equilibrium, while I have two different settings: a mixed-strategy equilibrium and a pure-strategy equilibrium. A further departure is that Amaldoss and Jain work with different roles assigned to different subjects in the experiment: once a subject was a low-reward or a high-reward subject, the role remained unchanged throughout the entire experiment; in my dissertation, each subject assumed different roles – each subject had equal numbers of treatments in which she was a leader and a follower;
- Nieken and Sliwka (2010) also theoretically and experimentally test an asymmetric tournament in which two agents have to decide between a safe and a risky strategy. As I stated in the previous chapter, Nieken and Sliwka also found three Nash equilibria that are similar to those I found in my theoretical analysis of this chapter: i) the leader chooses the safe strategy and the follower the risky strategy; ii) they both choose the risky strategy; iii) they both play a

mixed-strategy equilibrium. Differently, in my model there are only two possible equilibria, corresponding to equilibria i) and iii) of Nieken and Sliwka. Another different assumption by Nieken and Sliwka was to vary the correlations between the risky strategies – these correlations assumed values of 0, 1 and 1/2, respectively. In Nieken and Sliwka's initial model, the performance indicators for the risky strategies are correlated with a correlation coefficient $0 \leq \rho \leq 1$. Such a coefficient does not exist in my experimental model from the current chapter, because in my model the players' strategies are completely unobserved and simultaneous. In Nalebuff's initial sailing model, the correlation between strategies was 1 in the sense that once the leader left the finishing line the follower could imitate or adapt to the leader's strategy. In my adapted model, the correlation between the risky strategies of the two players is 0, in the sense that no player can see and copy the other's strategy. This zero correlation was in fact endorsed by Nalebuff himself, because sailing and R&D races are essentially different through the information structure: in sailing, one can see and copy the other's strategy; this is not possible with R&D races, because strategies are simultaneous and invisible. In other words, with R&D races it is not necessary to correlate the risky outcomes, because the strategies cannot be known, due to secrecy about the R&D strategy (to reinforce this point, see Tables 1.1 and 1.2 in Chapter 1 of this dissertation). However, Nieken and Sliwka varied this correlation to test how their equilibria change according to the correlation coefficient, and this was one of the main contributions from their paper. Therefore, my own contribution to the literature is important because I assume zero correlation between the risky strategies and I let vary other parameters which were not present in Nieken and Sliwka's paper, or in other papers: I let vary the probability of success for the risky strategy throughout the entire possible interval [0;1] and I also introduce a time framework through the accumulation of interest rates, parameters which I present in Chapter 6.1.2. One should notice that I only use one parameter, which I will denote as p, the probability of success for the risky project, and p is the same for both players. However, this does not mean that there is any correlation between the players' strategies, it only means that the players embark on similar R&D projects (with similar rates of success), and therefore one can assume, in order to avoid extreme mathematical and experimental complexity, that these risky R&D projects have the same chance to succeed (expressed through parameter p);

- in Bronars' (1986) tournament, the front runner prefers a low risk strategy to protect her position, while an opponent who is behind in the competition is motivated to choose a riskier strategy in order to catch up. Similarly, Knoeber and Thurman (1994) find that players with a higher ability in a tournament will choose less risky strategies than players with a disadvantage; I found similar results in my experimental study;
- Breitmoser et al. (2010) point out, just as I do in this dissertation, that the predictions of theoretical models are highly sensitive to small changes in the model parameters (Breitmoser et al. are referring to the predictions of perpetual race models developed by Aoki, 1991 and Hörner, 2004). However, when faced with an experimental setting, subjects' strategies are far less sensitive. With patent races, people are simpler and more stable than economic models. I do find similar results and draw conclusions on somewhat similar lines in my experimental study.

For all these reasons, Nalebuff's "simple" game-theoretical model has an appreciable depth which makes it suitable for mathematical and experimental testing. It pays respect to many critical assumptions of the patent race literature. At the same time it simplifies these assumptions and grasps the essence of patent race models in a way that would not be confusing for participants in an experiment. I therefore adopt Nalebuff's puzzle and build a patent race model upon it in the next chapter.

6.1.2. Building a Novel Patent Race Model

As in Nalebuff (1988), I look at a model of two firms entrenched in R&D competition with the goal of achieving a patent for a technology they are both developing. Only the first firm to reach successful development is awarded the patent. Firm 1 is six months ahead of Firm 2. I rename Nalebuff's strategy RISKY to strategy FAST, and denote strategy SAFE as strategy SLOW. The reason for this is that Nalebuff's original denomination might confuse participants in our later experiment. In fact, Nalebuff's strategy SAFE only ensures that the technology is completed; it does not imply that the completing firm actually receives the patent and thus it is not "safe". We now have strategy SLOW taking two years but leading to successful completion of the technology with probability 100%, and strategy FAST taking one year but leading to completion with varying probability p (instead of Nalebuff's fixed initial value of 50%). The prize for receiving the patent and capturing the market is denoted as x. Nalebuff does not model costs in his puzzle. However, since R&D competitions always imply significant investments, and costs are modeled in virtually every paper from the patent race literature, I introduced a cost parameter c. For reasons of simplicity, any strategy (SLOW or FAST) has the same cost c. This may sound like an over-simplification; however, it is reasonable to assume that the costs of a slower but secure technology development and of a faster but less secure development are not drastically different. Besides, having varying costs for the two strategies would make both our mathematical model and the experimental implementation overly complex. Moreover, a very similar assumption was used by Park (1987), who assumed that development costs are equal for a slow and safe strategy, as well as for a "jump" strategy (a risky strategy) with a small success rate.

A further parameter not included in Nalebuff's design which I nonetheless regard as highly relevant is time discounting. Time starts at $t = 0$ for both firms, where $t = 0$ is the moment in which Firm 1 decides on the strategy it wants to pursue. Measuring time in years, we have $t = 0.5$ as the moment when Firm 2 makes its own decision about the strategy it will follow. As we can see below in Figure 6.1, the latest moment in time in which any activity can still unfold is $t = 3$. I therefore apply an accumulation factor A reflecting the influence of the interest rate r as a function of time t. I denote $A = 1 + r$ as the annual accumulation, and $A^t = (1 + r)^t$ as the accumulation at point t in time. The accumulation is applied on all incurred costs and on the prize; the liquidation date is $t = 3$. Time point $t = 0$ marks the beginning of the game, and we consider both firms to have opportunity costs starting from this initial moment $t = 0$. This is based on the quite realistic assumption that both firms have money in their bank accounts, and both want to invest in an R&D project. I will later define the mathematical conditions under which the incentive to participate in the race exists for both firms. This participation constraint is essential for the existence and stability of a Nash equilibrium.

Figure 6.1. depicts the generalized decision tree. It contains all the possible outcomes of this game, depending on the players' decisions, on nature's moves, and on time. As in Nalebuff (and in the real world), each firm's move is independent and unobservable; secrecy about the other's move is of outmost importance. The symbols and are used to show the disruption of activities by Firm 1 and Firm 2, either because any of them won the patent, or because it lost the race.

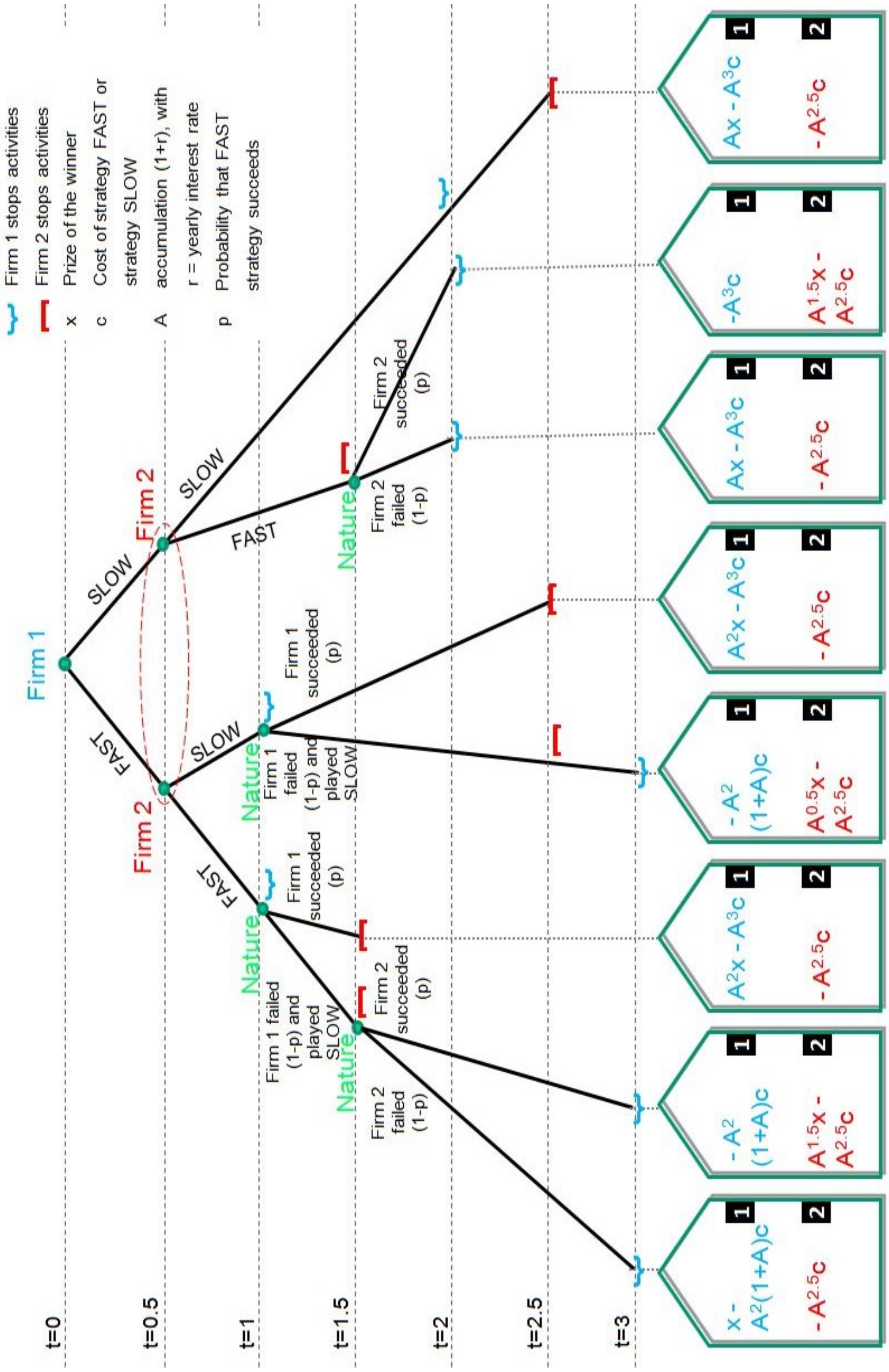
According to Nalebuff's racing puzzle, if a firm chooses strategy FAST, and fails, it is obliged to return to the beginning of the race, and choose strategy SLOW. However, in my model this can only be practically applied for Firm 1. The reason for this is that Firm 2 knows it has failed entirely if it fails with strategy FAST: no matter what it does, it has no more time to catch up with Firm 1. Therefore, Firm 2 only pays the cost of its strategy (its R&D cost) once. However, if Firm 1 fails with its FAST strategy at $t = 1$, it has no information about Firm 2's activities at that point in time. Thus, it is obliged to return to the beginning, choose strategy SLOW, and pay again its development cost.

I describe the game according to Rasmusen's (2001) definitions. This game has imperfect information (not every information set is a singleton), it is a game of uncertainty (Nature moves after the players move), it is symmetric[8] (no player has information different from the other players when she moves, or at the end nodes) and complete[9] (nature does not move first, or its initial move is observed by every player).

[8] With simultaneous decisions, no player has any informational advantages (Rasmusen, 2001).

[9] Two kinds of games have complete but imperfect information: games with simultaneous moves, and games where, late in the game, Nature makes moves not immediately revealed to all players (idem).

Figure 6.1. Generalized Decision Tree



In the following, I give the definition intervals for the parameters in the model, and provide reasoning for their choice.

x	the prize for the winner of the race, defined in the interval (0; 1]
c	the cost of any strategy, restricted to the interval [0; αx), where α is between [0; 1]
p	the probability of success for the FAST strategy, between [0; 1]. The most reasonable values for p can be found away from the boundaries of the interval definition
A	the accumulation factor, equal to 1 + r. I let the interest rate per annum r vary between the extreme values of [0%; 100%]. Therefore, A will vary between [1; 2]

In the choice of the definition intervals, I let the parameters assume very extreme values. This is in order for our equilibrium analysis to be complete and cover all possible ranges of parameter values. The cost c is originally restricted only to being lower than the prize x. The probability of success for the FAST strategy, p, takes all possible values between zero and one; however, the extreme case $p = 0$ makes strategy FAST ineligible (leading to an equilibrium (SLOW, SLOW), in which Firm 1 always wins due to its time advantage of six months), while the extreme case $p = 1$ eliminates the stochastic element of the model (leading to an equilibrium (FAST, FAST) in which strategy SLOW is always dominated by strategy FAST). I therefore do not expect relevant equilibria (especially under a participation constraint) to be found anywhere near the interval boundaries [0; 1]. The choice of the accumulation factor A directly depends on the way I defined r, the annual interest rate. In investment and financing theory (Schmidt and Terberger, 1997, p. 199-206), r expresses the same concept under different formulations: a) r it is the rate of return which stockholders claim as yield on their capital investment because they could achieve it as alternative return in a different application of funds; here, r is the alternative rate of return of the best achievable and comparable alternative; b) r is the minimal rate of return of a self-financed investment which protects the best interest of the stockholder; c) r is the discounting interest rate of the stock market; it is the interest rate at which actual and potential stock holders discount the advantages that are derived from the ownership of a share; d) r is the growth factor of a dividend D_t at time t, where $D_t = D_0 (1 + r)^t$. The main difference between the last two interpretations is that c) is viewed from time $t = 0$

looking into the future, while d) is viewed from time t looking into the past. For the experimental purposes of this chapter I will use interpretation d), because participating subjects will be paid in time $t = 3$. As stated before, I let r vary between 0% (reflecting a financial market with infinitely patient investors) and 100% (a market with very impatient investors).

I denoted $\pi_i^{s_1,s_2}$ as the payoff of Firm i for a strategy set (s_1, s_2), where (s_1, s_2) ϵ {(SLOW, SLOW); (SLOW, FAST); (FAST, SLOW); (FAST, FAST)}.

By following each ramification of the generalized decision tree in Figure 6.1, one can compute the expected payoff for each player i and each decision set (s_1, s_2).

For the strategy set (SLOW, SLOW), the expected payoffs are:

$$\pi_1^{S,S} = Ax - A^3c \tag{6.1}$$

$$\pi_2^{S,S} = -A^{2.5}c \tag{6.2}$$

For the strategy set (SLOW, FAST), the expected payoffs are:

$$\pi_1^{S,F} = -pA^3c + (1-p)(Ax - A^3c) \tag{6.3}$$

$$\pi_2^{S,F} = p(A^{1.5}x - A^{2.5}c) - (1-p)A^{2.5}c \tag{6.4}$$

For the strategy set (FAST, SLOW), the expected payoffs are:

$$\pi_1^{F,S} = p(A^2x - A^3c) - (1-p)(A^2 + A^3)c \tag{6.5}$$

$$\pi_2^{F,S} = -pA^{2.5}c + (1-p)(A^{0.5}x - A^{2.5}c) \tag{6.6}$$

For the strategy set (FAST, FAST), the expected payoffs are:

$$\pi_1^{F,F} = p(A^2x - A^3c) - (1-p)p(A^2 + A^3)c + $$
$$+ (1-p)(1-p)(x - (A^2 + A^3)c) \tag{6.7}$$

$$\pi_2^{F,F} = -pA^{2.5}c + (1-p)p(A^{1.5}x - A^{2.5}c) - $$
$$- (1-p)(1-p)A^{2.5}c \tag{6.8}$$

After rearranging the equation terms according to the prize *x* and the cost *c*, I obtained the payoff matrix of the game, which is depicted in Table 6.1.

Table 6.1. Payoff matrix of the game

		Firm 2	
		SLOW	FAST
Firm 1	SLOW	1 $Ax - A^3c$	1 $A(1-p)x - A^3c$
		2 $-A^{2.5}c$	2 $A^{1.5}px - A^{2.5}c$
	FAST	1 $A^2px + A^2(p-A-1)c$	1 $[pA^2 + (1-p)^2]x + A^2(p-1-A)c$
		2 $A^{0.5}(1-p)x - A^{2.5}c$	2 $A^{1.5}p(1-p)x - A^{2.5}c$

6.1.3. Equilibrium Analysis

Since each player has two strategies (SLOW or FAST) in her strategy set, there are only four strategy combinations that can be tested in the search for a Nash equilibrium in pure strategies. First of all, I show why (SLOW, SLOW) and (FAST, SLOW) can never be Nash equilibria.

(SLOW, SLOW) is a Nash equilibrium (Nash, 1951) in pure strategies if none of the players wants to deviate unilaterally from strategy set (SLOW, SLOW). In our game, this happens when the following conditions simultaneously hold: $\pi_1^{S,S} > \pi_1^{F,S}$ and $\pi_2^{S,S} > \pi_2^{S,F}$. The conditions can be translated into the following system of equations:

$$\begin{cases} Ax - A^3c > A^2\mathrm{px} + \mathrm{A}^2(p - A - 1)c & \textbf{(6.9)} \\ -A^{2.5}c > A^{1.5}px - A^{2.5}c & \textbf{(6.10)} \end{cases}$$

which easily simplifies to

$$\begin{cases} (1 - Ap)x + (1 - p)Ac > 0 & \textbf{(6.11)} \\ A^{1.5}px < 0 & \textbf{(6.12)} \end{cases}$$

From equation (6.12) it becomes clear that (SLOW, SLOW) can never be a Nash equilibrium, because $A^{1.5}px$ is never negative. In other words, Firm 2 would always deviate from the (SLOW, SLOW) strategy combination and choose the FAST strategy.

Analogously, we can test each other strategy combination for the existence of a Nash equilibrium. (FAST, SLOW) is a Nash equilibrium in pure strategies if $\pi_1^{F,S} > \pi_1^{S,S}$ and $\pi_2^{F,S} > \pi_2^{F,F}$ simultaneously hold, which translates to:

$$\begin{cases} A^2px + A^2(p - A - 1)c > Ax - A^3c & \textbf{(6.13)} \\ A^{0.5}(1 - p)x - A^{2.5}c > A^{1.5}p(1 - p)x - A^{2.5}c & \textbf{(6.14)} \end{cases}$$

and eventually simplifies as

$$\begin{cases} A(Ap-1)x + A^2(p-1)c > 0 & \textbf{(6.15)} \\ Ap-1 < 0 & \textbf{(6.16)} \end{cases}$$

Equations (6.15) and (6.16) can never hold at the same time. From (6.16) we have that $Ap - 1 < 0$, which implies that $A\,(Ap - 1)\,x < 0$ in equation (6.15). Furthermore, by definition $p - 1 \leq 0$, and thus $A^2(p - 1)\,c \leq 0$. Since the sum of a strictly negative number and a non-positive number cannot be a strictly positive number, equation (6.15) is contradicted by (6.16).

Therefore, the strategy combination (FAST, SLOW) is never a Nash equilibrium.

CASE 1

(SLOW, FAST) as a Candidate For a Pure-Strategy Nash Equilibrium

The problem of finding a set of parameters that will enforce (SLOW, FAST) as a pure-strategy Nash equilibrium must be broken down into two parts. First, we must prove the existence of an equilibrium from which no player wants to deviate unilaterally. Second, we must verify the participation constraints that will ensure that players actually have an incentive to enter the race (the existence of profit).

(SLOW, FAST) will be an equilibrium if $\pi_1^{S,F} > \pi_1^{F,F}$ and $\pi_2^{S,F} > \pi_2^{S,S}$, or:

$$\begin{cases} (1-A-p)px + A^2(1-p)c > 0 & \textbf{(6.17)} \\ A^{1.5}px > 0 & \textbf{(6.18)} \end{cases}$$

Furthermore, a player i will only enter the race if her payoff is larger than that of staying out (consistent with Rasmusen, 2001, p. 166). In this model, I normalized the outside option to zero. That is to say, if a player does not enter the race, she has nothing to lose, but also nothing to gain from the race, so her payoff for staying out is zero. The participation constraint for a player i is that $\pi_i^{S,F} > 0$, which leads to the following two equations for the two players:

$$\begin{cases} A(1-p)x - A^3c > 0 & \textbf{(6.19)} \\ A^{1.5}px - A^{2.5}c > 0 & \textbf{(6.20)} \end{cases}$$

Since there is no closed-form solution to the system of equations (6.17)-(6.20), I need to employ a graphical solution using the definition intervals for parameters x, c, A and p. I need to find out which sets of parameters x, c, A and p will simultaneously satisfy the four equations above and will therefore lead (SLOW, FAST) to become a Nash equilibrium in pure strategies. I will do so in the next section.

CASE 2

(FAST, FAST) as a Candidate For a Pure-Strategy Nash Equilibrium

According to the same reasoning as in CASE 1, I first test for the existence of a Nash equilibrium. (FAST, FAST) will be an equilibrium if $\pi_1^{F,F} > \pi_1^{S,F}$ and $\pi_2^{F,F} > \pi_2^{F,S}$. This is equivalent to:

$$\begin{cases} (Ap - A + 1 + p^2)x - A^2(1-p)c > 0 & \textbf{(6.21)} \\ p > \dfrac{1}{A} & \textbf{(6.22)} \end{cases}$$

As in CASE 1, the participation constraint that makes sure that any player i enters the race is that $\pi_i^{F,F} > 0$, and so:

$$\begin{cases} [pA^2 + (1-p)^2]x + A^2(p-1-A)c > 0 & \textbf{(6.23)} \\ A^{1.5}p(1-p)x - A^{2.5}c > 0 & \textbf{(6.24)} \end{cases}$$

As before, I will search for parameters x, c, A and p that simultaneously satisfy equations (6.21)-(6.24) through the use of a graphical solution.

CASE 3

Nash Equilibrium in mixed strategies

In a game lacking any pure-strategy equilibrium, there is always a mixed-strategy Nash equilibrium (Nash, 1951). The concept of a mixed strategy has found good application in a variety of contexts such as price promotion models (Varian, 1980), R&D modeling (Fudenberg and Tirole, 1985), bundled pricing (Lan and Kanafani, 1993), capacity choice (Deneckere and Peck, 1995), as well as product standardization (Farrell and Saloner, 1998). According to Harsanyi (1973), any mixed strategy m_i of player i will be of the form $m_i = \sum_k p_i^k s_i^k$, where p_i^k is the probability that the mixed strategy m_i assigns to the pure strategy s_i^k.

I denote in the game:

m_1 the probability that player 1 plays strategy FAST

m_2 the probability that player 2 plays strategy FAST

In order to compute these probabilities, one first needs to find the condition under which player 1 is indifferent between strategies FAST and SLOW. That is, one needs to know when player 1's expected payoff EP_1 is equal for any of the strategies. Sure enough, the condition will be a function of m_2, the probability for selection of strategy FAST of the other player - player 2. This translates into:

$$EP_1 = (1- m_2)\,\pi_1^{S,S} + m_2\,\pi_1^{S,F} = (1- m_2)\,\pi_1^{F,S} + m_2\,\pi_1^{F,F} \qquad \textbf{(6.25)}$$

from which we can derive the probability that player 2 chooses FAST:

$$m_2 = \frac{\pi_1^{S,S} - \pi_1^{F,S}}{\pi_1^{S,S} - \pi_1^{F,S} + \pi_1^{F,F} - \pi_1^{S,F}} \qquad \textbf{(6.26)}$$

$$m_2 = \frac{A(1-Ap)x + A^2(1-p)c}{[(1-p)^2 + Ap]x} \qquad \textbf{(6.27)}$$

Analogously, player 2 is indifferent between FAST and SLOW if her expected payoff EP_2 is the same for any strategy, or:

$$EP_2 = (1-m_1)\,\pi_2^{S,S} + m_1\,\pi_2^{F,S} = (1-m_1)\,\pi_2^{S,F} + m_1\,\pi_2^{F,F} \qquad \textbf{(6.28)}$$

which yields the probability that player 1 chooses FAST:

$$m_1 = \frac{\pi_2^{S,S} - \pi_2^{S,F}}{\pi_2^{S,S} - \pi_2^{S,F} + \pi_2^{F,F} - \pi_2^{F,S}} \qquad \textbf{(6.29)}$$

$$m_1 = \frac{Ap}{1 - p + Ap^2} \tag{6.30}$$

Interestingly, m_2 varies with all parameters x, c, A and p, while m_1 only varies with A and p. Differently, the probability with which player 1 chooses the FAST strategy is insensitive to variations of the prize x or the cost of any strategy c. The intuition behind this fact is that player 1's strategy will be more stable than player 2's strategy, as the former depends on fewer parameters. Indeed, the experimental findings will confirm this intuition.

Just as it is the case with the pure-strategy equilibria, we will need participation constraints to ensure that both players have an incentive to enter the race. This happens if their expected payoffs EP_1 and EP_2 are both positive. Using (6.25), (6.27), (6.28) and (6.30) I derive the following set of equations that need to be satisfied simultaneously in order for the mixed-strategy Nash equilibrium to be stable against the outside option:

$$\begin{cases} EP_1 = Ax - A^3c - \dfrac{A^2x^2p(1-Ap)}{\left[(1-p)^2 + Ap\right]x} - \\ - \dfrac{A^3(1-p)pxc}{\left[(1-p)^2 + Ap\right]x} > 0 & (6.31) \\ \\ EP_2 = -A^{2.5}c + \dfrac{A^{1.5}p(1-p)x}{1-p+Ap^2} > 0 & (6.32) \end{cases}$$

6.1.4. A Multi-Dimensional Graphical Solution

In this chapter I will plot the pure and mixed-strategy equilibria under varying parameter configurations x, c, A and p. Based on graphical depictions of the equilibria, this chapter will show how robust the equilibria will be to changes in the parameters. In the next chapter I will then select plausible sets of parameters that can later be used in a laboratory experiment.

The intuition behind this is the following belief, which I have expressed before: traditional theoretical patent race models, such as the ones described in Chapter 5.1, have failed to be confirmed experimentally because of their stiff parameters and narrow equilibria. These models are too complicated and inflexible for the human mind to grasp, or the human intuition to feel. I draw once more on Breitmoser et al. (2010), who said that traditional patent race models are too sensitive to parameter changes, while subjects' decisions are less sensitive. Other experimental studies which developed their own intuitive theoretical models (those described in Chapter 5.2.2) obtained excellent results which reunited theory and experiment.

In the model presented in this dissertation, I will seek to find the boundaries of each equilibrium depending on varying parameter sets, and to pick parameters that would be easy to process for human participants in an experiment. Furthermore, the parameters I pick will have to be as far away as possible from the equilibrium boundaries, and as deep as possible inside the equilibrium chart. The reason for this is that behavior should become unstable around the equilibrium boundaries, and more stable away from them. Say I selected two sets of parameters that are both close to the equilibrium boundaries, where there is a clash of the equilibria (and therefore a clash of the best strategies). In such points, only a machine can compute the right equilibrium and the right strategy. But with widely defined equilibrium paths, and parameter sets that are representative and help distinguish between the equilibria, one might expect humans to do a good job as well in choosing the payoff-maximizing strategy.

So I solved this five-dimensional model by constructing three parameter sets with a varying relationship between x and c. First of all, I normalized the prize to $x = 1$. In the experiment, this could be 1,000,000 Taler, which carries an emotional value for human participants of the type "Who wants to be a millionaire?". I let c vary from very small to unacceptably large. In the following, c will assume three values: 0.05, 0.25 and 0.45. In the experiment, the cost c of any strategy would then be: 50,000 Taler, 250,000 Taler and 450,000 Taler.

PARAMETER SET 1

Low Cost (x = 1, c = 0.05)

The following graphical representations have been programmed using the software Wolfram Mathematica 9.0.

Figures 6.2. and 6.3. show graphically the equilibrium charts for the Nash equilibria in pure strategies (SLOW, FAST) and (FAST, FAST) depending on *p* and *A*. Figure 6.4. shows the equilibrium chart for the Nash equilibrium in mixed strategies (m_1, m_2).

Figure 6.2. (SLOW, FAST) is a pure-strategy equilibrium only for the areas where all three graphs are visible simultaneously.

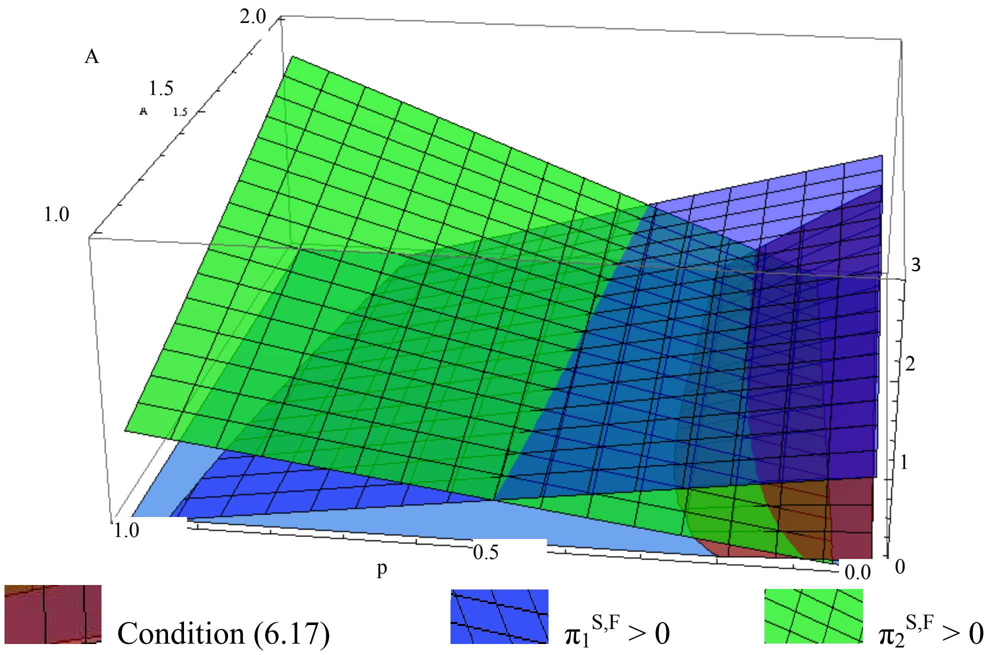


Figure 6.3. (FAST, FAST) is a pure-strategy equilibrium only for the areas where all three graphs are visible simultaneously.

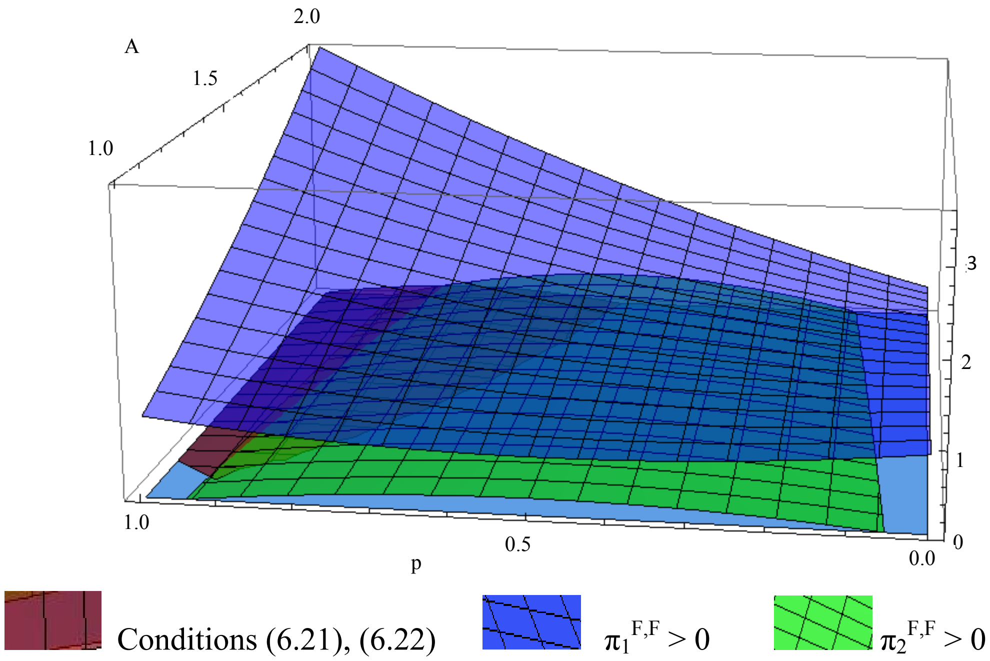
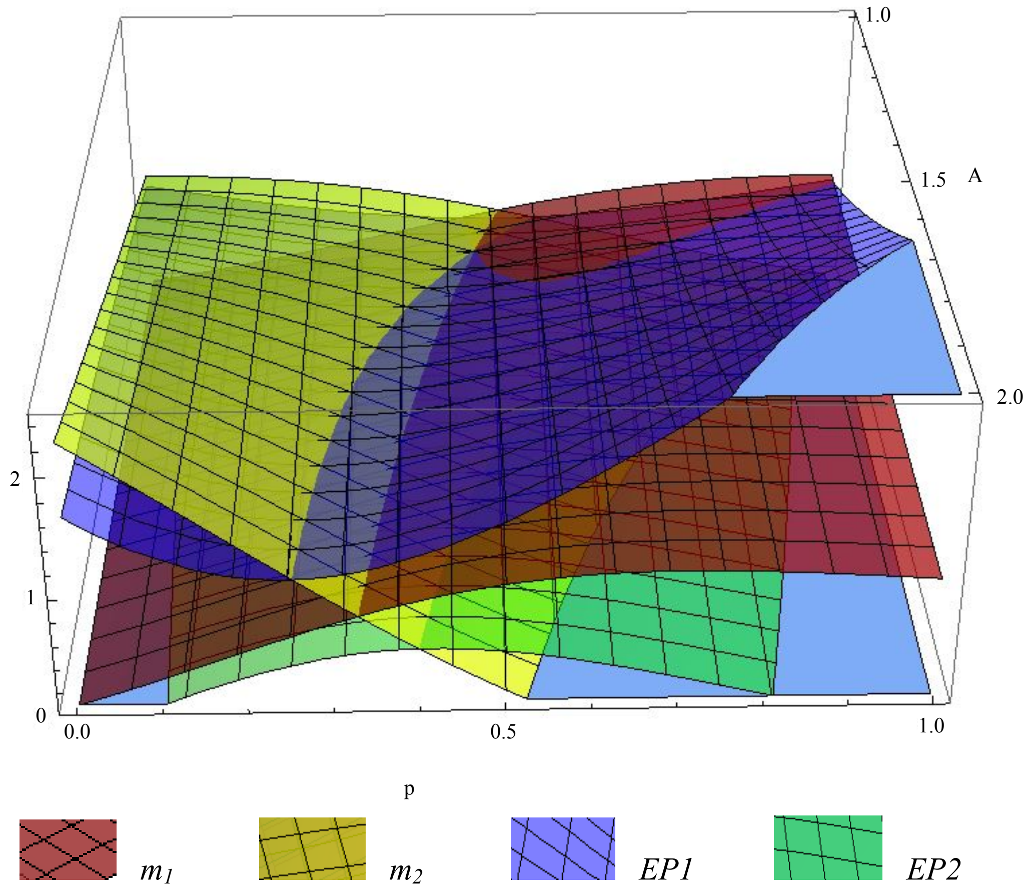
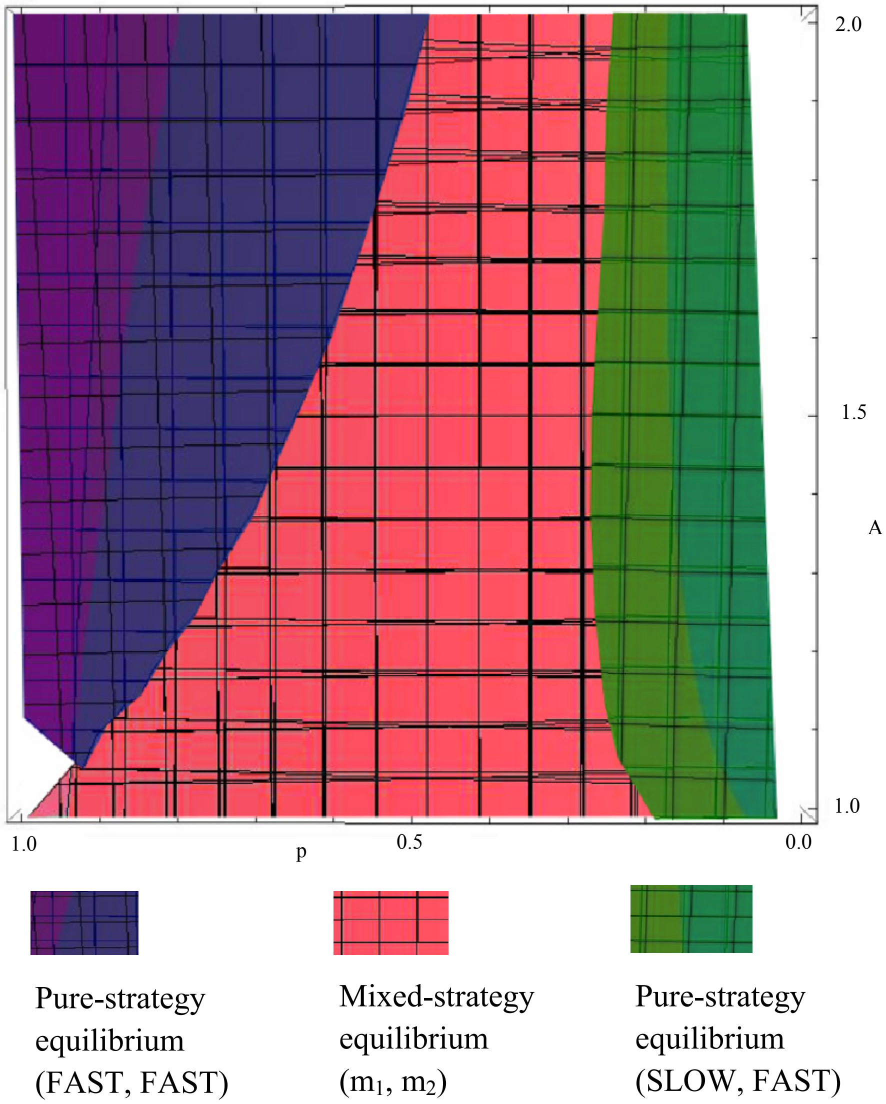


Figure 6.4. (m_1, m_2) is a mixed-strategy equilibrium only for the areas where all four graphs are visible simultaneously.



By rotating the three graphical representations from Figures 6.2, 6.3, 6.4 in Wolfram Mathematica, I could obtain a very precise graphical depiction for the boundaries of each of the three equilibria. That is, one can know under which parameter configurations p and A each equilibrium will begin and cease to exist. This is done in Figure 6.5, which combines and flattens all the three 3-dimensional graphs cases seen above on to a 2-dimensional plot varying only on p and A. In other words, one can see the projection of the 3-dimensional equilibria on the horizontal plane defined by p and A.

Figure 6.5. The three equilibria: (SLOW, FAST); (FAST, FAST) and (m_1, m_2) as bi-dimensional areas depending on *p* and *A*.



The boundary condition which separates equilibrium (FAST, FAST) from (m_1, m_2) is:

$$p = \frac{1}{A} \tag{6.33}$$

The boundary condition which separates equilibrium (m_1, m_2) from (SLOW, FAST) is:

$$(1 - A - p)p + 0.05A^2(1 - p) = 0 \tag{6.34}$$

PARAMETER SET 2

Medium Cost (x = 1, c = 0.25)

Similarly to Parameter Set 2, Figures 6.6, 6.7 and 6.8 show the equilibrium charts for the Nash equilibria in pure and mixed strategies (SLOW, FAST), (FAST, FAST) and (m_1, m_2) as functions of p and A.

Notice from Figure 6.7 that now the pure-strategy equilibrium (FAST, FAST) no longer exists, because there are only two graphs left: the red one representing the combined conditions (6.21) and (6.22), and the blue one showing that Firm 1 still finds it profitable to enter the race. However, the green graph has disappeared, showing that the participation constraint $\pi_2^{F,F} > 0$ is no longer satisfied, so there is no incentive left for Firm 2 to enter the race. The increase in the cost of R&D from 0.05 to 0.25 has made this equilibrium too unstable.

Figure 6.6. (SLOW, FAST) is a pure-strategy equilibrium only for the areas where all three graphs are visible simultaneously.

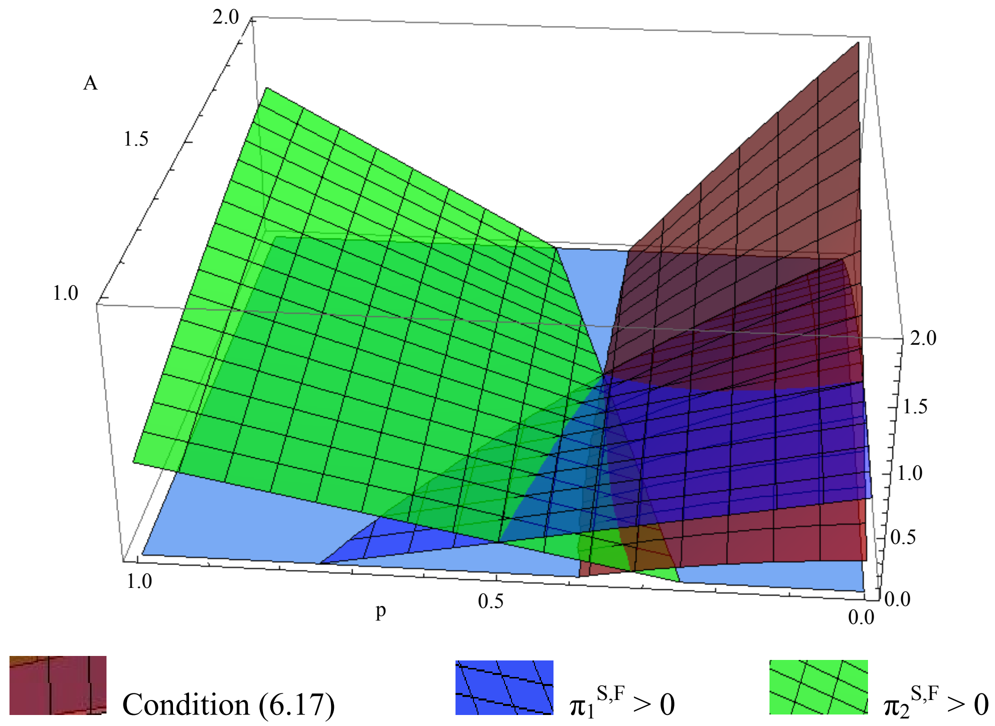


Figure 6.7. (FAST, FAST) is a pure-strategy equilibrium only for the areas where all three graphs are visible simultaneously.

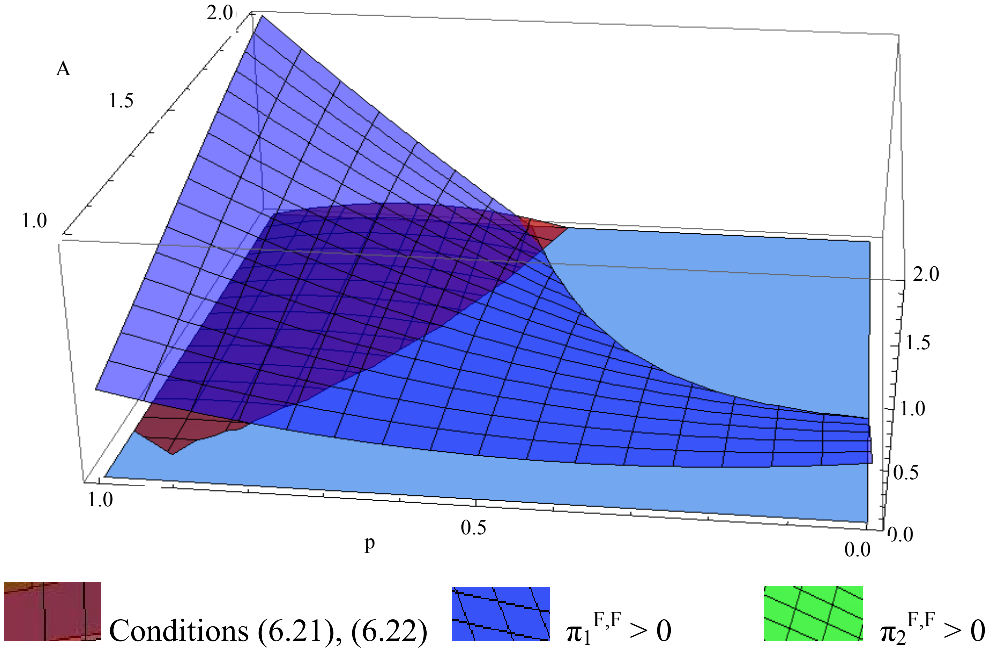


Figure 6.8. (m_1, m_2) is a mixed-strategy equilibrium only for the areas where all four graphs are visible simultaneously.

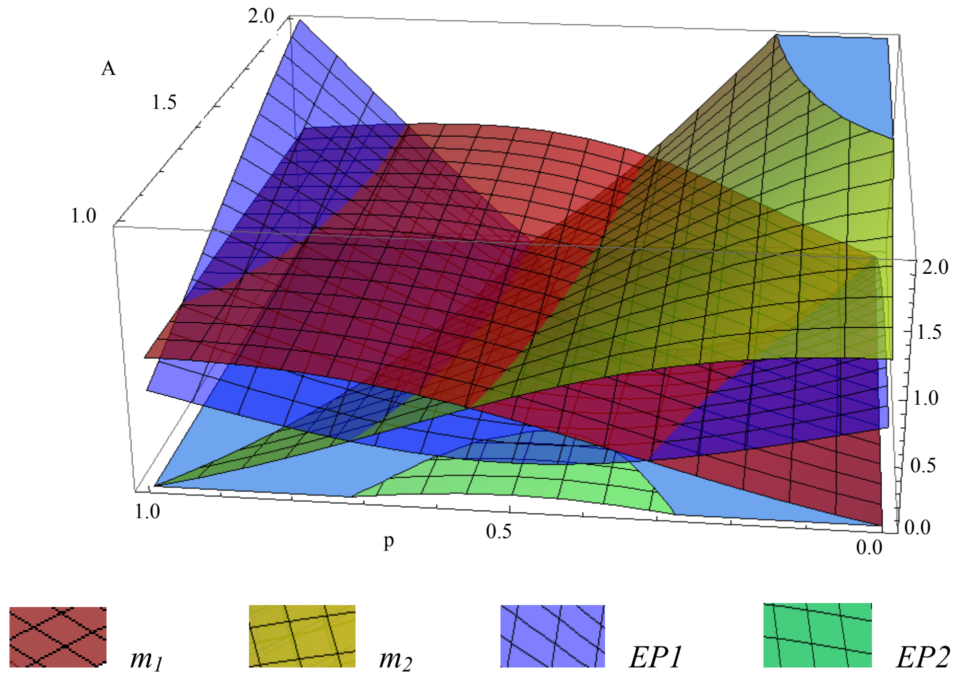
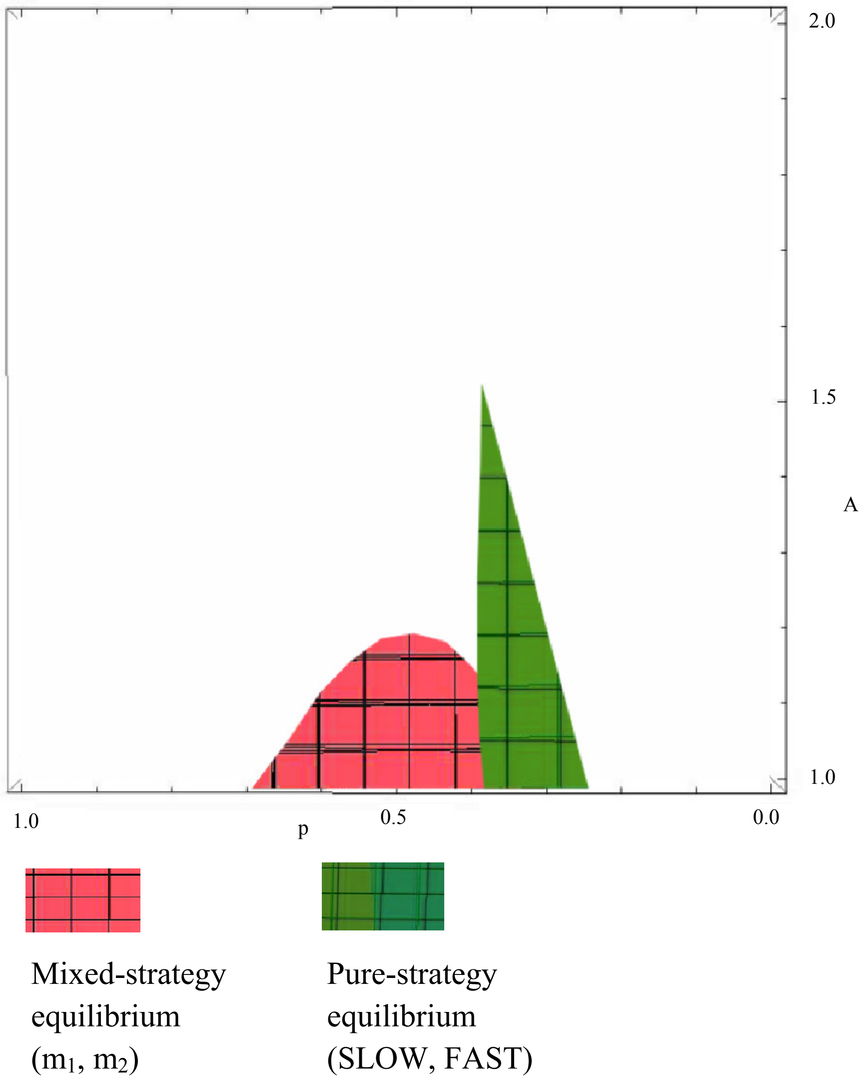


Figure 6.9. The two remaining equilibria: (SLOW, FAST) and (m_1, m_2) as bi-dimensional areas depending on *p* and *A*.



The boundary condition which separates equilibrium (m_1, m_2) from (SLOW, FAST) is:

$$(1 - A - p)p + 0.25A^2(1 - p) = 0 \qquad \textbf{(6.35)}$$

PARAMETER SET 3

High Cost (x = 1, c = 0.45)

Figure 6.10. (SLOW, FAST) is a pure-strategy equilibrium only for the areas where all three graphs are visible simultaneously.

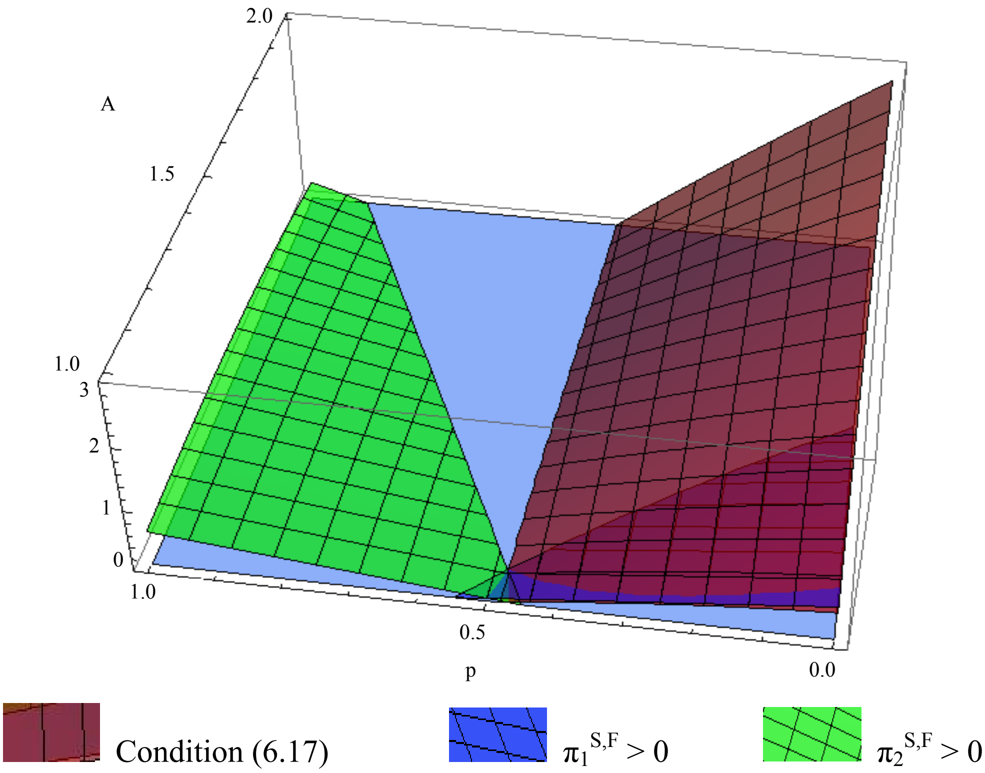


Figure 6.11. (m_1, m_2) is a mixed-strategy equilibrium only for the areas where all four graphs are visible simultaneously. No such area exists.

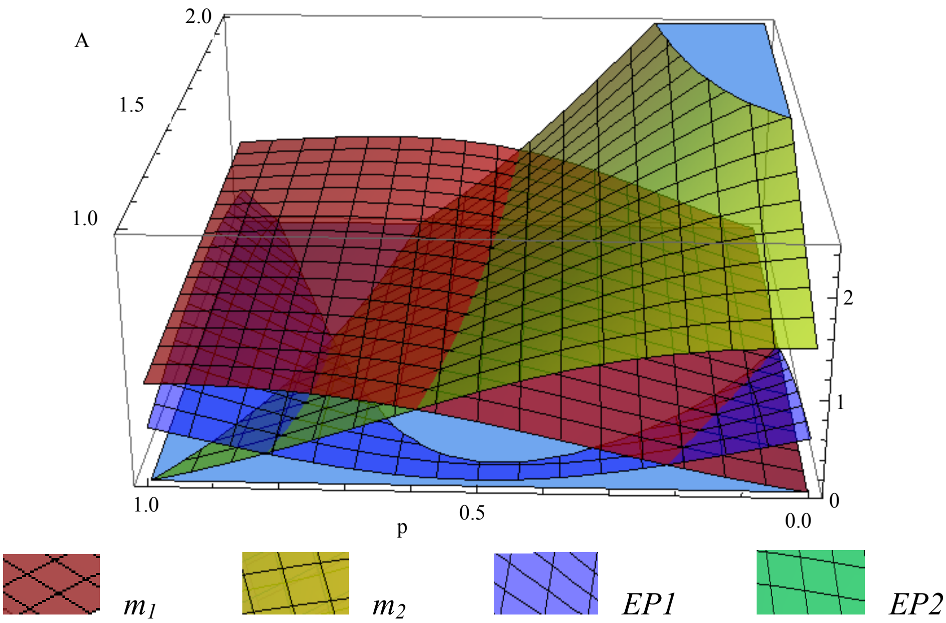
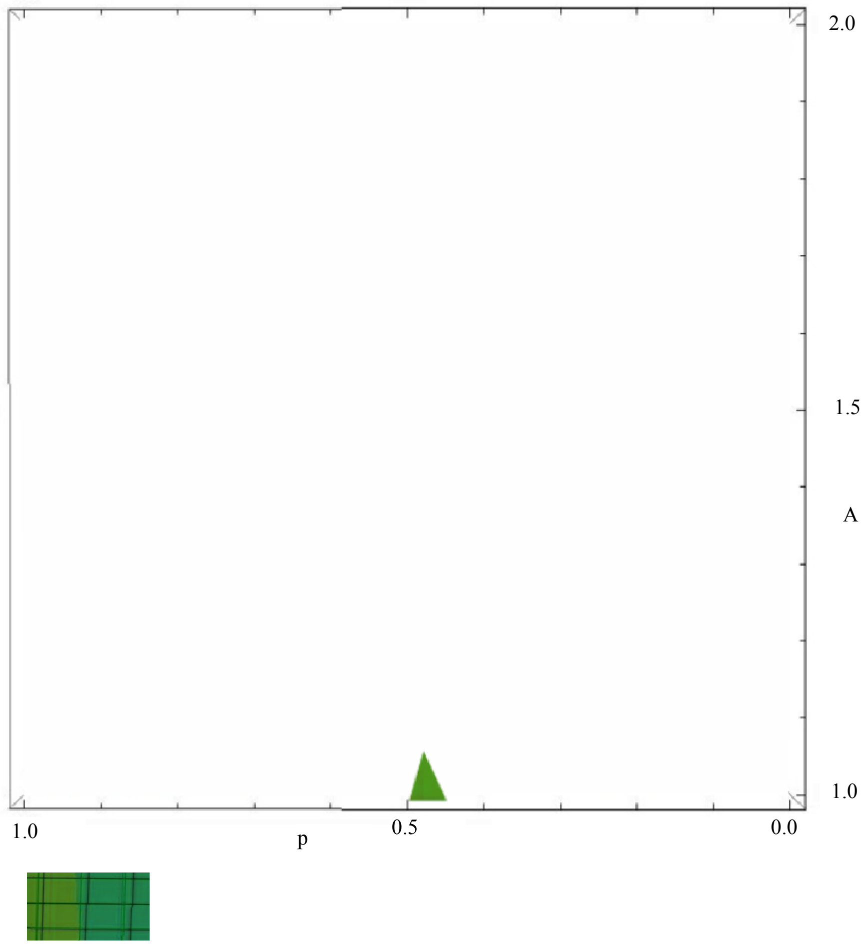


Figure 6.12. The only remaining equilibrium: (SLOW, FAST) as bi-dimensional area depending on p and A.



Pure-strategy
equilibrium
(SLOW, FAST)

6.1.5. Parameter Selection

I believe that the selection of parameters for a successful experimental implementation must take into account the following considerations:

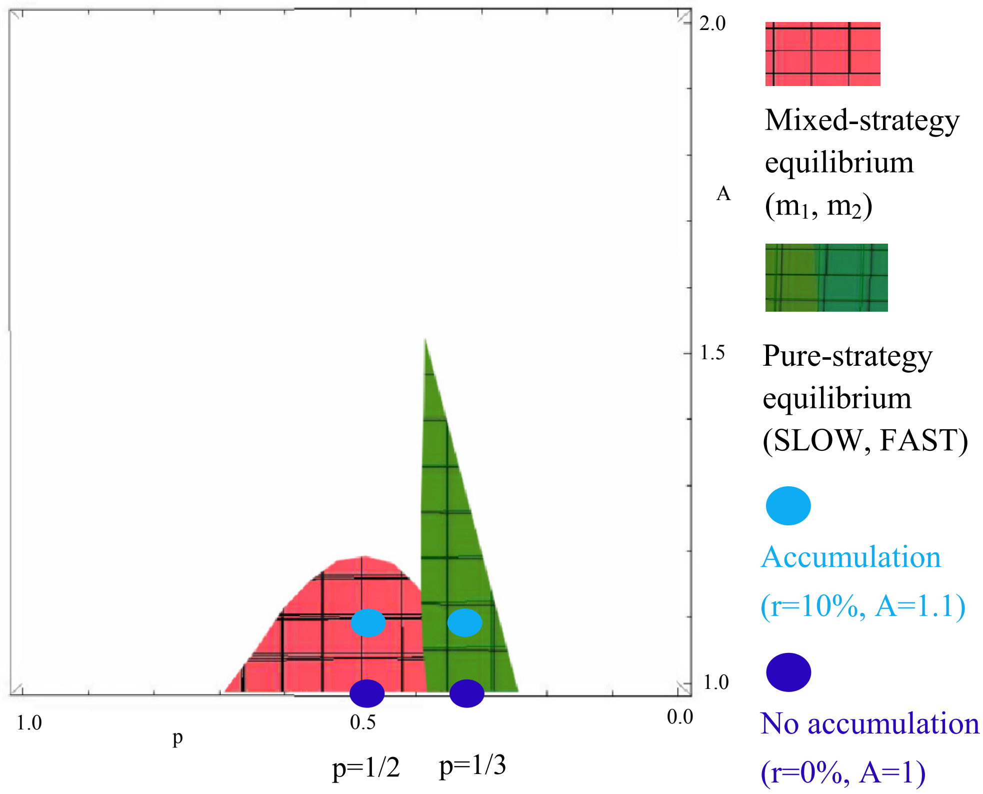
- the accumulation factor A should not be too large. Values between [1; 1.25] are recommended, corresponding to an interest rate r of up to 25% per annum. I took this high rate not because it could ever be a reasonable central bank interest rate in any stable economy. But I consider it the opportunity cost of other profitable and patentable foregone R&D projects (on the lines of Schmidt and Terberger, 1997, which I have discussed in Chapter 6.1.2). For example, Firm 2, which finds itself at disadvantage in the patent race, may choose to produce something altogether different, and thus the high interest rate represents the opportunity cost of the outside option;

- the recommended values for A between [1; 1.25] imply that we can, for the sake of choosing realistic parameters, rule out the (FAST, FAST) equilibrium altogether. This equilibrium needs to have the fulfilled condition $p > \frac{1}{A}$ (equation 6.22), and that would need $p > 0.8$, a probability too high to keep enabled the stochastic component of the game. Furthermore, I have an additional reason to eliminate the (FAST, FAST) equilibrium. One can see that it only exists for a very low cost of $c = 0.05$ in Parameter Set 1, and it fails to exist for a medium cost of $c = 0.25$ in Parameter Set 2. I consider that in Parameter Set 1, the cost-prize ratio is too low to express R&D investments. Usually, in patent races the R&D costs are also quite significant relative to the size of the prize. For example, I calculated the average ratio of cost to the prize in the experiment by Zizzo (2002) based on his experimental results: average investment per subject and round was 4.06 units; with a quadratic cost function and 16.1 rounds played by an average subject I obtained a total mean cost of $4.06^2 \times 16.1 = 265.38$ units, with a prize of 1000 units. This results in a cost-prize ratio of about 0.26, which is strikingly close to the cost-prize ratio generated by my medium cost parameter $c = 0.25$ (provided that only one strategy is played, and not both);
- one can see in Figure 6.12 that the introduction of a high cost $c = 0.45$ from Parameter Set 3 almost completely destroys all equilibria (by violating the participation constraint). The remaining trace of the pure-strategy equilibrium (SLOW, FAST) is so dismal and unstable that it would make little sense to formulate normative predictions based on it and test them experimentally.

Due to the above-mentioned reasons, I chose Parameter Set 2 with a medium-sized cost of $c = 0.25$ and a prize of $x = 1$. The next required step was to decide on the size of p and A. I looked at Figure 6.9 and took into account the considerations mentioned at the beginning of Chapter 6.1.4: that participants in an experiment should be faced with simple parameters and that those parameters should be situated well inside the equilibrium chart (not close to the equilibrium boundaries). From the myriad of options for p, two values stood out: a value of p = 1/2 would generate a mixed-strategy equilibrium, a value of p = 1/3 would generate a pure-strategy equilibrium (SLOW, FAST); both values are very easy to understand by participants and they are quite well in the middle of the equilibrium chart, or away from the boundaries. I employed the same criteria for the selection of values for the accumulation factor $A = 1 + r$. Participants would surely cope well with an interest rate of r = 0% and to one of r = 10% (meaning A = 1 or A = 1.1). Again, these values are stable inside the equilibrium chart.

My final selection of parameters is visible in Figure 6.13: I will test a mixed-strategy equilibrium versus a pure-strategy equilibrium and vary the accumulation factor. These four data points constitute the basis of the experimental design, which I describe in Chapter 6.2.2.

Figure 6.13. Parameter selection for p and A, given c = 0.25 and x = 1



6.2. Building the Patent Race Experiment

6.2.1. Experiments as Research Method

Laboratory experimental economics have been an important part of economic research for more than half a century. Laboratory experiments have been used, among others, to "investigate resource allocation issues in oligopolistic markets, public goods provision, and various auction market institutions" (Isaac and Reynolds, 1992). In laboratory experiments, real people make economic decisions for money and the quality of their decisions influence people's payoffs. "Behavioral rules for decision-making are not imposed ex ante, but are whatever flesh and blood human beings choose to do" (Isaac and Reynolds, 1992).

One of the building blocks of experimental economics is the induced value theory introduced by Smith (1976). He writes that "since economic theories always deal with certain alleged behavioral tendencies in isolation, the experimental laboratory is uniquely well suited for testing the validity of such theories" (idem). Smith also points out that all "characteristics of real world behavior" (e.g. "self-interest", "interdependent tastes, risk aversion") "arise naturally, indeed inevitably, in experimental settings" (idem). According to his induced value theory, control can be obtained in experimental studies by "using a reward structure to induce prescribed monetary value on actions" (idem), which was later called "incentive compatibility". Smith (1976) developed three principles that ensure experimental control: i) monotonicity (a higher payoff is preferred to a lower one); ii) salience (subjects' awareness that the quality of their decisions will influence payoffs); iii) dominance (nothing is more important for the respondent's utility than the experimental payoffs).

Schade (2005) compares field studies with economic experiments and points out that "dynamics are threatening the reliability of the different kinds of field studies; at the same time only performance dependent payments used in economic experiments may lead respondents to react sensitively to tradeoffs and incentives". A similar idea is expressed by Ariely (2008, p. xxi): "Why experiments? Life is complex, with multiple forces simultaneously exerting their influences on us, and this complexity makes it difficult to figure out exactly how each of these forces shapes our behavior.

For social scientists, experiments are like microscopes or strobe lights. They help us slow human behavior to a frame-by-frame narration of events, isolate individual forces, and examine those forces carefully and in more detail. They let us test directly and unambiguously what makes us tick".

In other words, field research in dynamic environments encounters the drawback that most relationships between analyzed variables change very rapidly. Therefore, field research "faces the challenge to discriminate relatively stable factors from the 'noise' of the couple of dozens of other rapidly changing factors", while economic experiments "are one potential means to overcome this 'dynamism obstacle'" (Schade, 2005). In such experiments, 'noise' and changing influences are reduced, the dynamic reality is not directly dealt with, the researcher obtains "control over the experimental situation" and "results may become replicable and generalizable to some extent" (idem). For these reasons, "economic experiments may also be used to compare competing theories" and they "help to modify existing and to build new theory" (idem).

Schade (2005) also compares the use of professionals versus students as subjects in experimental studies and notes that "experimental economists often judge it advantageous to experiment with students". He points out to studies in which professionals were not able to adapt to the experimental situation: they did not react to the opportunities and incentives in the experiment, but instead applied previous knowledge about the decision situation. Differently, students who were mostly inexperienced novices adapted better to the experimental incentives. Lévesque and Schade (2005) underline that experts, as opposed to novices, are more exposed to certain decision biases (e.g. overconfidence).

In the laboratory controlled, incentive-compatible experiment, the three principles of monotonicity, salience and dominance are fulfilled. My subject pool consists mostly of students, following Schade's (2005) recommendation that "the incentives and tradeoffs faced in the decision when to exploit a business opportunity may well be studied in an incentive-compatible experiment with students". Since my theoretical model is especially designed to analyze competitive behavior in a patent race model, I expect such behavior to shine out through the experiment with students.

6.2.2. Experimental Design and Implementation

At the end of Chapter 6.1.5, I selected two sets of parameters for p and A (generating four parameter combinations) and gave plausible reasons for this choice. I let the prize $x = 1$ and the cost of any strategy $c = 0.25$ stay constant during the experiment, and I transform them into Taler values with a psychological importance for participants: the prize x becomes 1,000,000 Talers and the cost c is equal to 250,000 Talers. The probability of success with strategy FAST, p, is varied between the two values 1/3 (reflecting the pure-strategy equilibrium SLOW, FAST) and 1/2 (leading to the mixed-strategy equilibrium m_1, m_2). The accumulation factor is varied between $A = 1$ ($r = 0\%$, no accumulation) and $A = 1.1$ ($r = 10\%$, accumulation). Furthermore, I want any subject who participates in the experiment to be confronted with all four parameter combinations, and to be confronted with the two possible roles: that of Leader and that of Follower in the race. The variation of the three variables (p, A and role) generated $2^3 = 8$ settings, which are depicted in Table 6.2.

Table 6.2. Eight experimental settings S1-S8

Setting	Role	Equilibrium	Probability of success for strategy FAST	Accumulation
Setting 1	Leader	Pure	p = 1/3	No Accumulation
Setting 2	Leader	Mixed	p = 1/2	No Accumulation
Setting 3	Leader	Pure	p = 1/3	Accumulation (r=10%)
Setting 4	Leader	Mixed	p = 1/2	Accumulation (r=10%)
Setting 5	Follower	Pure	p = 1/3	No Accumulation
Setting 6	Follower	Mixed	p = 1/2	No Accumulation
Setting 7	Follower	Pure	p = 1/3	Accumulation (r=10%)
Setting 8	Follower	Mixed	p = 1/2	Accumulation (r=10%)

In order to test if subjects follow the pure-strategy and especially the mixed-strategy equilibrium, I need each setting to be played over several rounds. I decided that a number of seven rounds for each of the eight settings S1-S8, totaling 56 rounds per subject, would be appropriate because it allows for the computation of both mixed-strategy and pure-strategy equilibria without making the experiment too long for participants. For example, if a subject decides to choose strategy FAST in four out of the seven rounds of Setting 2, one can state that the subject played a mixed-strategy equilibrium with a probability of choosing strategy FAST equal to 4 / 7 = 57.14 %. Similarly, if a subject chooses strategy FAST in none of the seven rounds of Setting 1, one can say that the subject played the pure-strategy equilibrium (SLOW, FAST).

The patent race experiment was programmed using the software z-Tree (Fischbacher, 2007) and was carried out in July 2012 in the experimental laboratory of the School of Business and Economics of the Humboldt-Universität zu Berlin. Subjects were mostly students (87,5%) from 23 different fields of study (n = 48 subjects, with 36 female and 12 male participants). I ran four sessions, each with twelve subjects matched into six pairs. Subjects were told that they were going to play a competition game against a fixed opponent. Learning was not possible since there was no feedback about the outcomes of previous rounds throughout the experiment. Each subject played 56 rounds and the order of the eight settings was randomly varied from one round to the next. Respondents were informed that 14 rounds (seven rounds as Leaders and seven rounds as Followers) would be randomly selected by the computer at the end of the experiment. The detailed written experimental instructions describing the decision situation and the parameters of the game were handed out to the participants in German. An English translation of the instructions is available in Appendix 6A of this dissertation. The instructions were read out by the experimenter and then presented again on the computer screen. Subjects had to correctly answer eight comprehension questions before they could make their 56 decisions. Afterwards, I measured subjects' risk attitudes using the procedure by Holt and Laury (2002) and I gathered their demographic information. The average duration for a session was of approximately 80 minutes. On average, 45 minutes were spent on making the actual 56 decisions. Subjects earned € 14.54 on average, with a minimum of € 11.40 and a maximum of € 20.20.

6.2.3. Propositions and Hypotheses

In this chapter I first use the normative game-theoretic model which was completed at the end of Chapter 6.1 in order to derive benchmark propositions about how subjects should play. These propositions show us what strategies a perfectly rational economic agent would choose, under the given parameters. This approach is based on assumption that individuals are able to formally optimize on economic variables.

Second, I formulate behavioral hypotheses, based on the framework of bounded rationality developed by Simon (1957). Bounded rationality is the idea that decision makers seek a satisfactory solution rather than the mathematically optimal solution, because their rationality is limited by the information they have, by cognitive gaps and by the finite amount of time available to them for making a decision. Simon coined this phenomenon "satisficing". Gigerenzer and Selten (2002, p.14) point out that there is more to Simon's approach than just "satisficing", and that aspiration levels (e.g. the aspiration of a decision maker to achieve a certain profit on a certain market) are an essential feature of bounded rationality. Decision makers have certain aspirations which are dynamically adjusted to the situation: aspirations are raised if it is easy to accomplish them, and lowered if satisfactory outcomes are hard to acquire. Todd and Gigerenzer (2003) find three interpretations for the concept of bounded rationality: i) "optimization under constraints"; ii) "erroneous deviations from a rational standard"; iii) "a fit between the structures of the mind and the environment" in the form of "simple, bounded, heuristics that fill up the mind's adaptive toolbox" (idem). Lévesque and Schade (2005) experimentally study intuitive optimization and point out the last two approaches ii) and iii) best account for the bounded rationality phenomenon. They suggest that humans make errors in some situations, but perform better than expected in others.

Based on the first approach, the *homo economicus* approach, which assumes perfect rationality, I formulated two benchmark propositions which were derived from the equilibrium predictions elicited in Chapter 6.1. Proposition 1 expresses the pure-strategy equilibrium SLOW, FAST. Proposition 2 implements the selected parameters to compute a mixed-strategy equilibrium. I used equations (6.27) and (6.30) from Chapter 6.1.3 to calculate the probabilities of play for strategy FAST, denoted as m_1 and m_2.

<u>Proposition 1</u>

"Given a low probability of success for strategy FAST p = 1/3
*the Leader will play FAST with probability **0%***
*and the Follower will play FAST with probability **100%**."*

<u>Proposition 2</u>

"Given a medium probability of success for strategy FAST p = 1/2
without accumulation (r = 0%)
*the Leader will play FAST with probability **66%***
*and the Follower will play FAST with probability **83%***
with accumulation (r = 10%)
*the Leader will play FAST with probability **71%***
*and the Follower will play FAST with probability **81%**."*

Based on the second approach, the approach of bounded rationality, I assumed that individuals would employ heuristics and simple dynamics in their decisions. I was able to hypothesize the application of two heuristics in my patent race experiment. The first of these heuristics, and the theoretical basis for Hypothesis 1, is the extremeness aversion by Tversky and Simonson (1992), according to which individuals have a tendency to stay away from alternatives that are described by extreme combinations of characteristics. A second possible behavioral pattern, the basis for Hypothesis 2, is represented by anchoring-and-adjustment heuristics (Slovic and Lichtenstein, 1971). Individuals select a certain initial value, called the anchor, and later adjust in the direction suggested by the decision situation. Tversky and Kahneman (1974) say that "different starting points yield different estimates, which are biased toward the initial values'' (p. 1128).

<u>Hypothesis 1</u>

(Extremeness aversion)

"If given the opportunity, individuals will mix between strategies, even if the rational choice is to play a pure strategy."

<u>Hypothesis 2</u>

(Anchoring-and-adjustment heuristics)

"Individuals' first decision situation will influence their subsequent decisions."

6.3. Econometric Analysis

In this chapter I compare the normative predictions derived from the equilibrium analysis with the actual decisions made by subjects in the experiment, and I employ statistical methods to test whether those predictions had a significant effect on subjects' choices. Furthermore, I analyze the distributions of subjects' deviations from equilibrium play and I provide explanations for these deviations. Finally, I verify whether subjects employed anchoring-and-adjustment heuristics and I describe why they did or did not. By running the econometric analysis, I also test the validity of the benchmark propositions and the hypotheses formulated in the previous chapter.

6.3.1. Normative Predictions and Subjects' Decisions

In the first part of the analysis I present descriptive statistics about subjects' predicted percentages of play for strategy FAST, their actual mean choices, their deviations from normative predictions and the significance levels of those deviations. This is done in Table 6.3.

Table 6.3. Comparing benchmark propositions with actual play

	Predicted mean	Observed mean	Deviation	p-value
for *p* = 1/3 *The Leader will play FAST with probability*	**0%**	***11%	11%	0.000
and the Follower will play FAST with probability	**100%**	***50%	-50%	0.000
for *p* = 1/2 *Without accumulation:* *the Leader will play FAST with probability*	**66%**	**50%	-16%	0.005
and the Follower will play FAST with probability	**83%**	75%	-8%	0.054
With accumulation: *the Leader will play FAST with probability*	**71%**	*59%	-12%	0.019
and the Follower will play FAST with probability	**81%**	74%	-7%	0.144

Significance levels: *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$

One can conclude from Table 6.3 that people behave surprisingly close to the equilibrium predictions. All deviations from optimal play, except one, are small (not larger than 16% in absolute values) and most of them are significant or highly significant. There is only one large and significant deviation, for the settings in which the Followers should have played a pure-strategy equilibrium but clearly played mixed strategies (the deviation being -50% from optimality).

In the next part of my analysis I run regressions in order to explain what factors influence the dependent variable Setting Decision. This variable expresses the percentage of FAST choices made by each subject in any of the settings S1-S8. It means that the variable includes 48 subjects x 8 settings = 384 data points. Also, the data points are correlated for each of the 48 subjects, and thus I need to employ a panel econometric model to estimate the influence of independent variables on the Setting Decision. I build hierarchical regressions, starting with one predictor and incorporating several others step by step. The essential predictor to start with is the normative prediction, which I called Setting Prediction.

When working with panel data, one has to find out whether a fixed-effects or a random-effects regression model is appropriate to estimate the data. Whenever time-invariant variables (e.g. age, gender, risk attitude) are included in the regression, the decision is quite easy: in fixed-effects models all time-invariant variables are omitted, and therefore a random-effects model should be employed. Even though in all my later regressions I will use time-invariant predictors, in the first regression I am using only Setting Decision as dependent variable and Setting Prediction as predictor. Since none of them is time-invariant (but instead they vary with time), the best way to decide between a fixed-effects and a random-effects model is to estimate both and to run a Hausman test which gives guidance on the better model.

I used the software Stata/IC 12 to produce all regressions provided in this chapter and in Appendices 6B-6C. Tables 6.4 and 6.5 show the estimators for the fixed-effects and the random-effects models, respectively. Table 6.6 contains the results of the Hausman test. One can see that the estimates of the coefficients and the standard errors of both models are almost equal. The Hausman test suggests that there are no systematic differences between the two models' coefficients. In order to keep the consistency with my later regressions, I select the random-effects model as the relevant one.

Table 6.4. Fixed-effects regression. Dependent variable: Setting Decision. Independent variable: Setting Prediction

```
Fixed-effects (within) regression               Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.3003                         Obs per group: min =         8
       between =      .                                        avg =       8.0
       overall = 0.2491                                        max =         8

                                                F(1,335)           =    143.81
corr(u_i, Xb)  = -0.0000                        Prob > F           =    0.0000
```

Setting Decision	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Setting Prediction	*** .5127595	.0427587	11.99	0.000	.42865	.5968689
Constant	*** .1575831	.0313294	5.03	0.000	.095956	.2192103
sigma_u	.16238716					
sigma_e	.31742843					
rho	.20742159	(fraction of variance due to u_i)				

```
F test that all u_i=0:     F(47, 335) =     2.09             Prob > F = 0.0001
```

Significance level: *** $p < 0.1\%$

Table 6.5. Random-effects regression. Dependent variable: Setting Decision. Independent variable: Setting Prediction

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.0000                         Obs per group: min =         8
       between = 0.0000                                        avg =       8.0
       overall = 0.2491                                        max =         8

                                                Wald chi2(1)       =    143.81
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

Setting Decision	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Setting Prediction	*** .5127595	.0427587	11.99	0.000	.4289539	.5965651
Constant	*** .1575831	.035616	4.42	0.000	.087777	.2273893
sigma_u	.11736477					
sigma_e	.31742843					
rho	.12026414	(fraction of variance due to u_i)				

Significance level: *** $p < 0.1\%$

Table 6.6. Hausman test comparing the fixed-effects and the random-effects model

	Coefficients (b) fixed	(B) random	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
Prediction	.5127595	.5127595	-2.44e-15	.

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

The main result of the selected random-effects model depicted in Table 6.5 is that the normative predictions did have a positive and highly significant effect on subjects' actual decisions in each setting. This result is also in line with that from Table 6.3, where we saw that the deviations from optimality are significant but mostly rather small. This means that people managed to intuitively behave in the direction of the equilibrium predictions, even though it was impossible for them to compute the optimal solution mathematically (as this would have implied them to do the full analysis of Chapter 6.1 in a matter of seconds). One can see in Table 6.5 that the factor Setting Prediction explained about 25% of the variation in Setting Choice (expressed through the overall R-squared).

For the sake of completion, I would like to explain what the values of the *F test* in the fixed-effects model, the *Wald chi* in the random-effects model, *sigma_u*, *sigma_e* and *rho* represent. The *F test* and the *Wald chi* are values which are compared to certain thresholds and determine the significance level of the regression. In order to understand the other concepts one has to know what a panel econometric model does. It equates the values of the dependent variable with the sum of the following parts: i) a constant term; ii) observed time-variant factors (here, the normative predictions); iii) observed time-invariant factors (here, age, gender, risk attitude); iv) an unobserved individual specific effect (e.g. preparation, intelligence, ability); v) an unobserved random error term (a residual). The *sigma_u* and *sigma_e* are estimates of the standard deviation of the individual specific effect and of the error term, respectively. *Rho* is the share of the estimated variance of the overall error accounted for by the individual specific effect.

I also wanted to test for the influence of other variables such as gender, age, the interaction between gender and role (Leader or Follower) as well as risk attitude on the dependent variable Setting Choice. In Appendix 6B, I built the corresponding hierarchical regressions and incorporated the variables step by step. None of these variables had any significant effect on the dependent variable. For the variable “risk attitude”, I had to eliminate about 25% of all data, because subjects violated the utility axioms from Holt and Laury (2002) and therefore their risk attitudes were inconsistent. I chose the most relevant regression, which is depicted in Table 6.7.

Table 6.7. Random-effects regression. Dependent variable: Setting Decision. Independent variables: Setting Prediction, Female, Age, Female * Role (the interaction of Female and Role)

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.3007                         Obs per group: min =         8
       between = 0.0312                                        avg =       8.0
       overall = 0.2547                                        max =         8

                                                Wald chi2(4)       =    145.06
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

Setting Decision	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Setting Prediction	*** .5274838	.05616	9.39	0.000	.4174122	.6375553
Female	.0343174	.060108	0.57	0.568	-.0834922	.152127
Age	.0067591	.0068704	0.98	0.325	-.0067066	.0202247
Female * Role	.0199047	.0491339	0.41	0.685	-.0763961	.1162054
Constant	-.0567872	.1891639	-0.30	0.764	-.4275415	.3139672
sigma_u	.11856147					
sigma_e	.3178252					
rho	.12215922	(fraction of variance due to u_i)				

Significance level: *** p < 0.1%. Data coding: Female = (1 if female, 0 if male); Role = (1 if Leader, 0 if Follower).

The main result from Table 6.7 reinforces the previous result from Table 6.5: the only significant factors influencing subjects’ decisions were the normative predictions, while all other variables had no significant effect. The percentage of variance explained remained almost the same around 25%. People’s behavior was therefore guided by equilibrium play.

6.3.2. Subjects' Deviations From Equilibrium Strategies

In this chapter I focus on finding explanations for subjects' deviations from equilibrium play. I do this by running new hierarchical regressions on the dependent variable "absolute deviation from equilibrium" and by analyzing the distribution of deviations in the experiment. First I must explain why I chose the absolute deviation and not the actual deviation. The Deviation is defined as the Setting Decision minus the Setting Prediction. The Absolute Deviation is the absolute value of Deviation. In half of the experiment (in all settings with $p = 1/3$, or pure-strategy equilibria SLOW, FAST), the Leader is constrained to having only positive deviations (because her equilibrium prediction is 0%), while the Follower is constrained to having only negative deviations (her equilibrium prediction is 100%). These constraints make the variable Deviation less appropriate to study in the regression. In order to have a clean account of which subjects deviated *at all* from the equilibrium (no matter in which direction), I used the dependent variable Absolute Deviation.

I ran a first regression with only one predictor, the variable Role (Leader or Follower). I made again the choice between a fixed-effects and a random-effects model by using the Hausman test, which suggested no systematic difference between coefficients. Like in the previous chapter, I therefore selected the random-effects model, which is depicted in Table 6.8. The fixed-effects model and the Hausman test can be found at the beginning of Appendix 6C. The predictor Role had a highly significant negative effect on the dependent variable Absolute Deviation ($p < 0.1\%$). This means that Followers deviated significantly more, and Leaders less from the equilibrium predictions. The amount of variance explained (the overall R-squared) was of 6.75%.

I then ran hierarchical regressions, incorporating step by step the following variables as predictors: the success probability of strategy FAST, the accumulation, gender, the interaction between gender and role, age and risk attitude. All these hierarchical regressions can be found in Appendix 6C. I selected the most relevant regression and represented it in Table 6.9. None of the newly incorporated predictors had any significant effect on the Absolute Deviation. The only significant predictor remained the variable Role; however, it appears that the inclusion of many new variables affected

Table 6.8. Random-effects regression. Dependent variable: Absolute Deviation. Independent variable: Role

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.0000                         Obs per group: min =         8
       between = 0.0000                                        avg =       8.0
       overall = 0.0675                                        max =         8

                                                Wald chi2(1)       =     29.06
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

Absolute Deviation	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Role ***	-.1532616	.0284305	-5.39	0.000	-.2089843	-.0975389
Constant ***	.3641232	.0221105	16.47	0.000	.3207874	.407459
sigma_u	.06377259					
sigma_e	.27856067					
rho	.04980155	(fraction of variance due to u_i)				

Significance level: *** p < 0.1%. Data coding: Role = (1 if Leader, 0 if Follower).

Table 6.9. Random-effects regression. Dependent variable: Absolute Deviation. Independent variables: Role, Success Probability of FAST, Accumulation, Female, Female * Role, Age

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.0865                         Obs per group: min =         8
       between = 0.0041                                        avg =       8.0
       overall = 0.0737                                        max =         8

                                                Wald chi2(6)       =     31.62
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

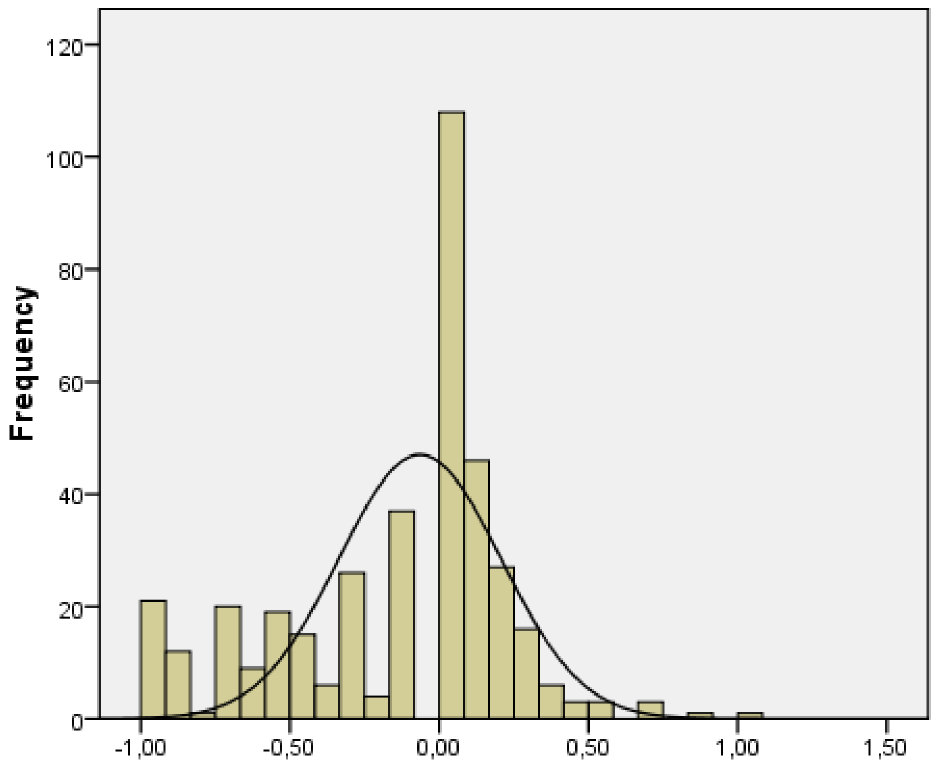
AbsoluteDeviation	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Role *	-.1295554	.0569095	-2.28	0.023	-.241096	-.0180147
SuccessProbabilityFAST	-.219734	.1707286	-1.29	0.198	-.5543559	.114888
Accumulation	.2091875	.2845477	0.74	0.462	-.3485157	.7668907
Female	-.0015233	.0519088	-0.03	0.977	-.1032627	.100216
Female_Role	-.0316083	.0657135	-0.48	0.631	-.1604044	.0971877
Age	-.0002575	.0050329	-0.05	0.959	-.0101217	.0096068
_cons	.2437238	.3371545	0.72	0.470	-.417087	.9045345
sigma_u	.06784575					
sigma_e	.27879867					
rho	.05590851	(fraction of variance due to u_i)				

Significance level: * p < 5%. Data coding: Role = (1 if Leader, 0 if Follower); Female = (1 if female, 0 if male).

the significance level of variable Role, which deteriorated from the 0.1% level in the first regression to the 5% level in the regression from Table 6.9. The R-squared only increased marginally from 6.75% to 7.37%.

The main conclusion from the analysis of the Absolute Deviation remains that Followers deviated more, and Leaders less, from normative predictions. To explain why, I look at the overall distribution of deviations for all subjects and all settings, and then separately for Leaders and Followers.

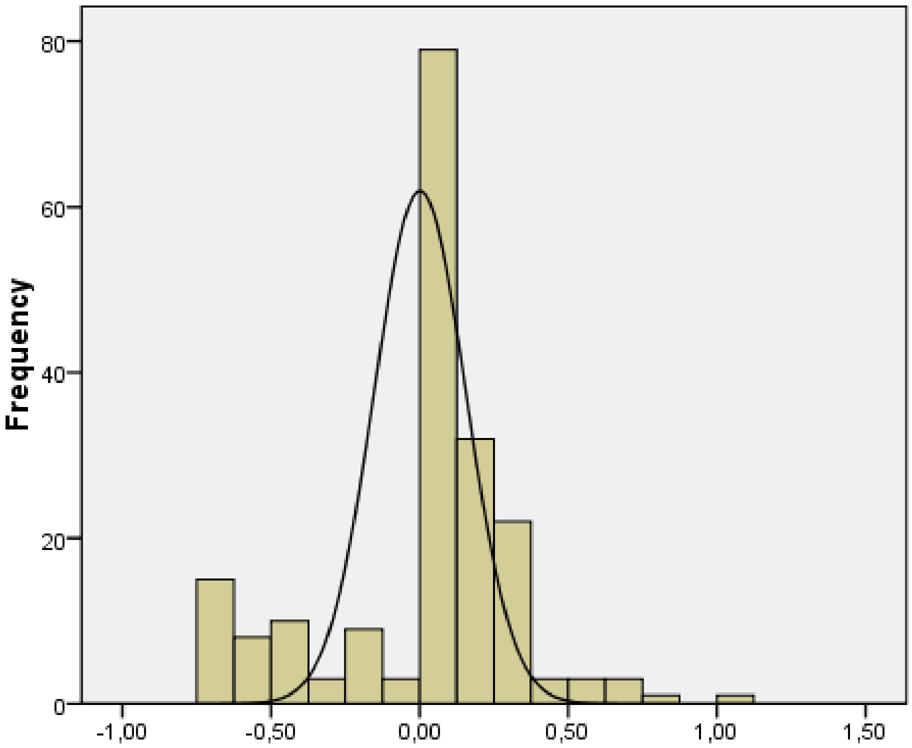
Figure 6.14. Distribution of deviations from normative predictions for all subjects and all settings



No. of observations	384
Mean	-0.154
Median	0
Mode	0
Standard deviation	0.383
Skewness	-0.581
Kurtosis	-0.041

Figure 6.14. shows the histogram of deviations from equilibrium predictions for all 48 subjects and all 8 settings, amounting to a total of 384 observations. Figures 6.15. and 6.16. show the distributions of deviations for Leaders and Followers, respectively. At first glance, it appears as if the deviations for all settings and for Leaders have a shape somewhat resembling the normal distribution. Not so for the deviations of Follower settings only, which more resemble the uniform distribution. The

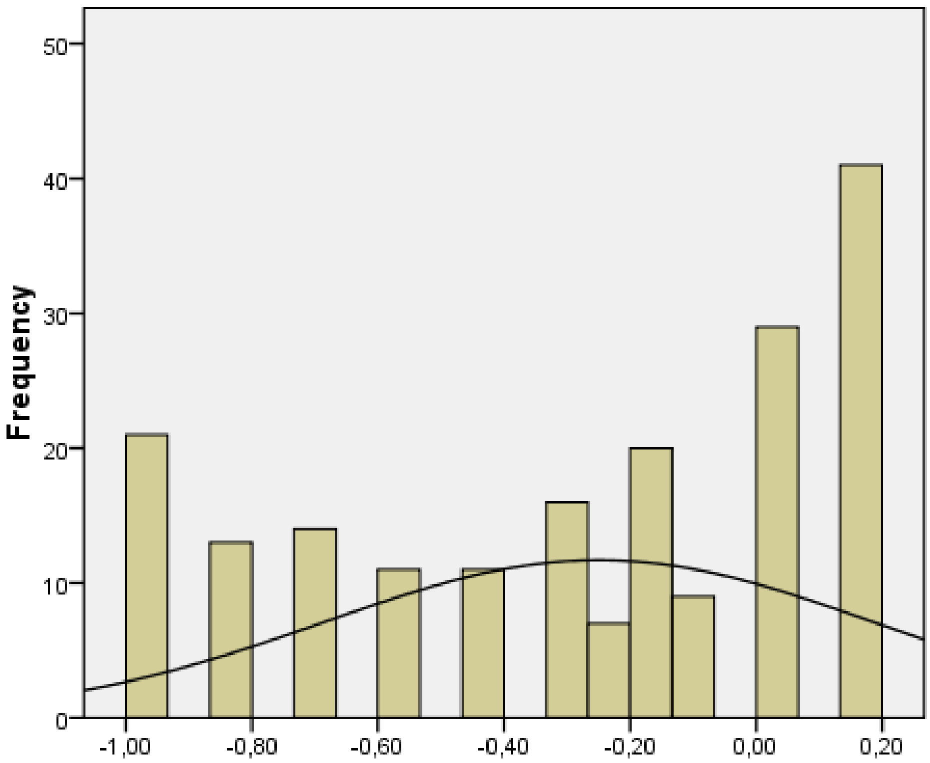
Figure 6.15. Distribution of deviations from normative predictions for all subjects as Leaders only



No. of observations	192
Mean	-0.013
Median	0
Mode	0
Standard deviation	0.317
Skewness	-0.322
Kurtosis	0.777

analysis of mean, median and mode all show that the smallest deviations from equilibrium predictions are achieved by Leaders (-0.013; 0 and 0, respectively) and large deviations are observed for Followers (-0.296; -0.142 and multiple modes of -1 and 0.14). Especially important here is the mode of zero for Leaders, with 31.77 % of all Leaders' decisions (427 out of 1,344 individual decisions, in 61 of 192 settings) made exactly according to the equilibrium prediction (SLOW, FAST).

Figure 6.16. Distribution of deviations from normative predictions for all subjects as Followers only



No. of observations	192
Mean	-0.296
Median	-0.142
Multiple modes	-1; 0.14
Standard deviation	0.393
Skewness	-0.537
Kurtosis	-1.050

The standard deviation of the distribution is lower for Leaders (0.317) than for Followers (0.393). The lower negative levels of kurtosis and skewness show a flatter distribution with larger tails for Followers[10]. I also ran Q-Q plots and computed Kolmogorov-Smirnov and Shapiro-Wilk tests of normality in SPSS, which generated no relevant additional information.

All indicators support the initial characterization of followers' distribution as resembling the uniform distribution. To sum up:

- Followers' deviations from equilibrium predictions took values which were more spread out and uniform, showing some uncertainty about what strategy to play;
- Leaders played much closer to equilibrium predictions and their deviations were exactly zero in about one third of all settings decisions.

Figures 6.17 and 6.18 may provide explanations about the differentiated deviation phenomenon. Figures 6.17 and 6.18 are simple re-runs of Figures 6.6 and 6.8 (from Chapter 6.1.4), but they also include the positions of the selected parameters (from Chapter 6.1.5) in the three-dimensional space generated by p, A and equilibrium-relevant variables defined in p and A. One can see in both figures that the equilibrium definition is more stable for Leaders and more fragile for Followers. In Figure 6.17, the two selected parameters (marked with red) are very close to the equilibrium boundary of the Follower, the line where the green graph (representing the Follower's payoff or $\pi_2^{S,F}$) meets the zero plane. On the other hand, the Leader's payoff or $\pi_1^{S,F}$, depicted by the blue graph, is visibly higher from the zero plane. I have also computed the actual levels for both payoffs by inserting the concrete values of parameters x, c, A and p into equations (6.3) and (6.4) from Chapter 6.1.2. We have the values $x = 1$ and $c = 0.25$ (Parameter Set 2 from Chapter 6.1.4), which remain unchanged for both Figures 6.17 and 6.18. In Figure 6.17, for the red point defined by parameters $A = 1$ and $p = 0.3333$ we find that $\pi_1^{S,F} = 0.4167$ and $\pi_2^{S,F} = 0.0833$ and therefore the Leader's payoff is much higher above zero than the Follower's payoff. For the other red point defined by parameters $A = 1.1$ and $p = 0.3333$ we find that $\pi_1^{S,F} = 0.4834$ and $\pi_2^{S,F} = 0.0672$, which confirms the result obtained with the previous parameters. In Figure 6.18 we find a similar situation in

[10] The interpretation of kurtosis and skewness is based on Field (2009, p. 138).

Figure 6.17. Selected parameters inside the (SLOW, FAST) pure-strategy equilibrium chart

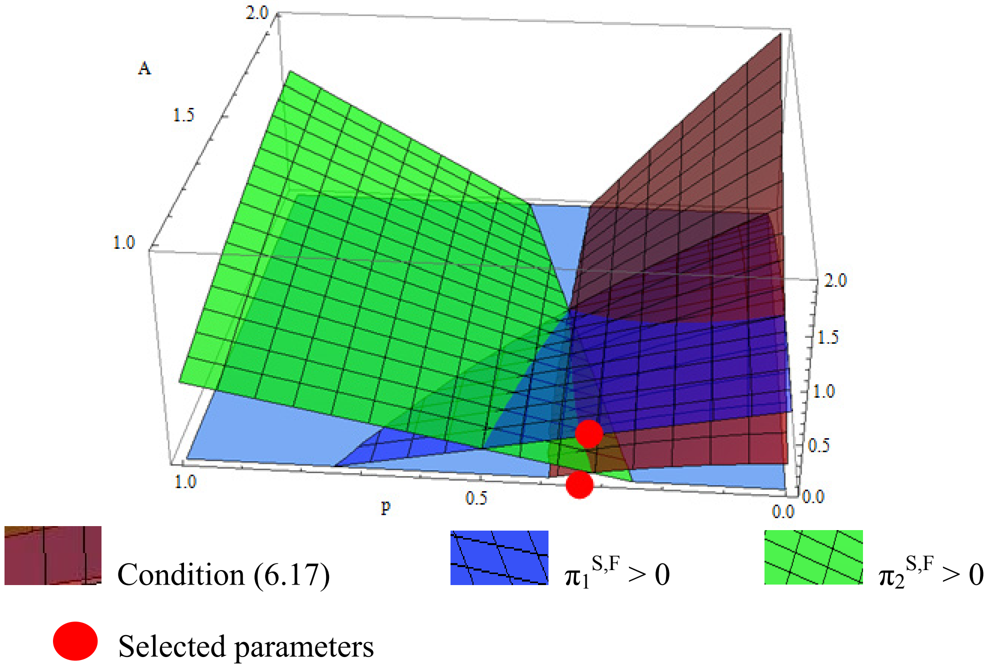
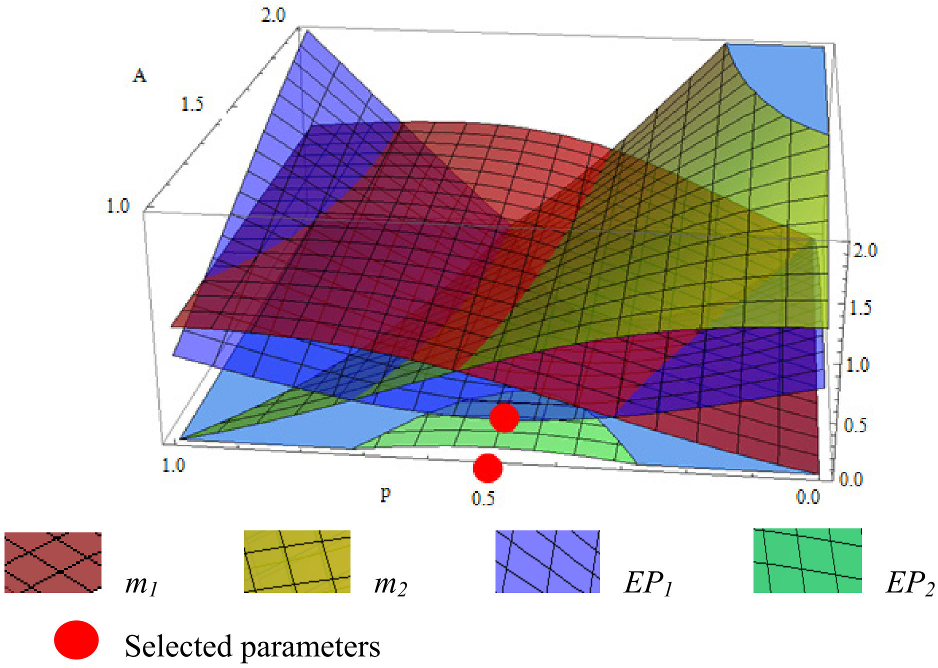


Figure 6.18. Selected parameters inside the (m_1, m_2) mixed-strategy equilibrium chart



that the green graph, the expected payoff of the Follower (EP_2) is visibly lower than the blue graph, the expected payoff of the Leader (EP_1), and EP_2 is much closer to the zero plane. To confirm this with precision, I have computed EP_1 and EP_2 with the given parameters x, c, A and p. EP_1 and EP_2 were computed in equations (6.31) and (6.32) in Chapter 6.1.3. For the red point defined in the zero plane by A = 1 and p = 0.5 one finds that EP_1 = 0.3333 and EP_2 = 0.0833. Similarly, for the red point defined by A = 1.1 and p = 0.5 the expected payoffs are EP_1 = 0.2474 and EP_2 = 0.0548. Thus, for all selected parameters in both Figures 6.17 and 6.18 the generated equilibria are less stable for the Follower than for the Leader.

To sum up the above findings: due to the asymmetric positions for Leaders and for Followers, which generate different expected payoffs – that is, lower payoffs for Followers which sometimes are just above zero levels – I find that equilibrium predictions are more unstable for Followers than for Leaders. I believe this explains the Followers' larger and more spread out deviations from equilibrium play that I have encountered in this chapter.

6.3.3. Testing Anchoring and Adjustment, Propositions and Hypotheses

In this chapter I test whether subjects employed anchoring-and-adjustment heuristics in their decisions, and I also provide the general results on the propositions and heuristics I formulated in Chapter 6.2.3.

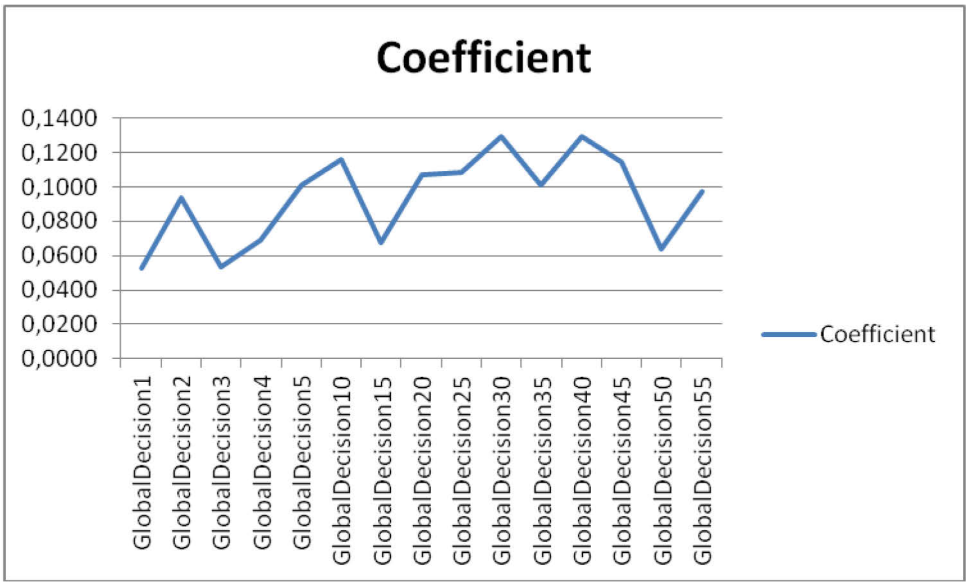
In order to check whether individuals' first decision situation (or first few situations) influenced their later decisions, I ran in Stata/IC 12 ordinary least squares regressions on the dependent variable "average decision" of subjects, having as independent variables the decisions made by subjects in the first five rounds and then in later rounds. I denoted the independent variables as "global decisions". Each variable contained a number of 48 observations from all subjects. Table 6.10 presents the coefficients, standard errors, significance values and amount of variance explained, while Figure 6.19 shows only the evolution of coefficients throughout the 56 rounds.

Table 6.10. Ordinary least squares regressions. Dependent variable: average decision. Independent variables: global decisions

Predictors	Coefficient	Standard Error	p-value	Adjusted R-squared
GlobalDecision1	0.0524	0.0483	0.284	0.0037
GlobalDecision2	*0.0937	0.0454	0.045	0.0649
GlobalDecision3	0.0537	0.0471	0.260	0.0063
GlobalDecision4	0.0694	0.0464	0.142	0.0256
GlobalDecision5	*0.1014	0.0451	0.030	0.0793
GlobalDecision10	*0.1161	0.0442	0.012	0.1115
GlobalDecision15	0.0673	0.0467	0.157	0.0223
GlobalDecision20	*0.1071	0.0447	0.021	0.0915
GlobalDecision25	*0.1086	0.0446	0.019	0.0950
GlobalDecision30	**0.1294	0.0436	0.005	0.1427
GlobalDecision35	*0.1011	0.0450	0.030	0.0792
GlobalDecision40	**0.1291	0.0441	0.005	0.1385
GlobalDecision45	*0.1146	0.0443	0.013	0.1081
GlobalDecision50	0.0640	0.0464	0.175	0.0188
GlobalDecision55	*0.0971	0.0452	0.037	0.0715

n = 48 observations. Significance levels: * p < 5%; ** p < 1%

Figure 6.19. Regression coefficients. Dependent variable: average decision. Independent variables: global decisions



The coefficients from Table 6.10 tend to become largest, with the highest levels of significance and largest percentage of variance explained, during rounds 20-45. This shows that there is an individual tendency within each subject to play strategy FAST (as evidenced by the average decision), but subjects learn about their own tendency during play. In the early rounds (0-20) participants' decisions are still far away from their average decisions, but they converge to the average decisions later in the experiment (rounds 20-45). Finally, during the last rounds (45-56) decisions diverge again from the average decision, probably due to boredom effects.

In any event, these results show that there is no evidence for the application of the anchoring-and-adjustment heuristic, because subjects' first decision situations did not significantly influence subsequent decisions. Thus, there is no support for Hypothesis 2. It appears that individuals' self-consistency is stronger than the anchoring phenomenon. Finally, Table 6.3 from Chapter 6.3.1 shows that the benchmark propositions 1 and 2 cannot be confirmed, because the *homo economicus* approach is a too hard assumption. However, it is true that individuals mixed between strategies even when the rational choice was to play a pure strategy (see Table 6.3) and therefore Hypothesis 1, based on the extremeness-aversion heuristic, is confirmed. Recall from Chapter 6.2.3 that the extremeness-aversion heuristic (Tversky and Simonson, 1992) was stating that individuals have a tendency to stay away from alternatives described by extreme characteristics. In my experiment, this means that subjects had a tendency to stay away from playing only the SLOW strategy or only the FAST strategy (which were the extreme choices of the subjects), even though the normative equilibrium predictions (in the form of the pure-strategy equilibrium) were showing that this was the purely rational choice that subjects should have followed. The results for the two propositions and the two hypotheses are summarized in Table 6.11.

Table 6.11. Results of the two propositions and the two hypotheses

Propositions and hypotheses	**Theoretical basis**	**Results**
Proposition 1	*Homo oeconomicus* approach	Not confirmed
Proposition 2	*Homo oeconomicus* approach	Not confirmed
Hypothesis 1	Extremeness-aversion heuristic	Confirmed
Hypothesis 2	Anchoring-and-adjustment heuristic	Not confirmed

6.4. And the Winner Is…? …The Human Being

Throughout this chapter, I developed a game-theoretic patent race model, tested it experimentally and analyzed it econometrically. In Chapter 6.1 I started with Nalebuff's (1988) sailing race model, transformed it into a game-theoretic model and added parameters such as cost and time discounting to adapt it to a patent race context. I then performed an equilibrium analysis and selected parameters that were easy to understand by participants and stable inside the equilibrium chart. The analysis predicted a pure-strategy and a mixed-strategy equilibrium. In Chapter 6.2 I shed light on the importance of experiments for research and I built an experimental design with eight settings by varying three parameters: i) the type of equilibrium; ii) the existence of accumulation through an interest rate; iii) whether a subject was at an advantage in the race (Leader) or at a disadvantage (Follower). I also formulated benchmark propositions based on the assumption of full rationality and hypotheses based on the concept of bounded rationality.

In Chapter 6.3 I presented descriptive statistics and performed regression analyses, which provided the experimental results. First of all, in Chapter 6.3.1 I found that experimental participants behaved suprisingly close to the equilibrium predictions, as their deviations from optimal play were mostly small. The results of a random-effects regression showed that the equilibrium predictions had a highly significant effect on subjects' actual choices. My conclusion from the regression was that individuals intuitively behave in the direction of optimal play, even though the experimental settings change randomly from round to round, and it is impossible for participants to compute the equilibrium mathematically. This brings support to the concept of intuitive optimizing by individuals (Lévesque and Schade, 2005). Indeed, in this experiment subjects made errors in some situations, but performed especially well in others.

Second, I wanted to find out why subjects deviated from equilibrium play in some situations, but decided very close to optimality in other situations. Chapter 6.3.2 contained random-effects regressions on the absolute deviation from equilibrium and found out that the only significant factor influencing deviations was a subject's role: when a subject was a Leader, she deviated significantly less from normative predictions than when a subject was a Follower. I analyzed the distribution of deviations for Leaders versus Followers and concluded that, indeed, Leaders' deviations were smaller and resembled more the normal distribution, while Followers deviated more and their decisions seemed to come from something similar to the uniform distribution. This implies that Followers encounter some uncertainty about what strategy to play (and therefore randomize between strategies). Overall, my experimental results seem to confirm the strength of game-theoretical modelling. The larger deviations from optimal play for Followers are likely to be the result of the asymmetric positions between the two roles and the less stable definition of the equilibrium for Followers. Whenever the Leader's payoff was rather high, the Follower's payoff was visibly lower and closer to the zero threshold under which the equilibrium predictions would have broken (see the end of Chapter 6.3.2 for details). Thus, the less stable equilibrium predictions for Followers seem to be reflected in less stable strategies and larger deviations from optimal play, which can be seen as a scientific result of game theory. Variables such as gender, age, accumulation and risk attitude had no significant effect on subjects' decisions or on their deviations.

Third, I ran ordinary regressions in Chapter 6.3.3 to test for anchoring and adjustment. I found that subjects had individual tendencies to play the strategy FAST rather than strategy SLOW, that they learned about their individual tendencies throughout the experiment, and that they did not employ anchoring-and-adjustment heuristics (showing that self-consistency was stronger than the anchoring phenomenon). The benchmark propositions derived from the perfect rationality approach were, as expected, not confirmed. However, I could find support for the extremeness-aversion heuristic, as subjects mixed between their strategies even when it was optimal for them to play pure strategies. Moreover, the extremeness aversion was more pronounced for Followers than for Leaders, because the equilibrium predictions from the game-theoretical model were less stable for Followers. In other words, Followers were less sure than Leaders that they should play pure strategies when the optimal play recommended them to do so, and therefore Followers avoided extreme combinations (such as the pure FAST strategy) and played more mixed strategies, showing some uncertainty about how they should act.

I consider the experimental study described in this dissertation to be an important contribution to the patent race literature in particular and to the study of competitive behavior in general. Experimental studies such as Zizzo (2002), Silipo (2005) and Breitmoser et al. (2010), which tested models from the theoretical patent race literature, found no support for this literature (see Chapter 5.2.1). As stated before, the reason for this was the fragility of such models, in which predictions change dramatically as a consequence of small changes in parameters or assumptions. Such models have the drawback that they were not based on proven and established concepts from economic theory, but instead on many parameters and assumptions. The situation was different for studies that constructed their own patent race models based on strong theoretical foundations and then tested them experimentally: Isaac and Reynolds (1992) proved Schumpeterian competition in an R&D race, Amaldoss and Jain (2002) showed the power of the concept of a mixed-strategy equilibrium in an asymmetric patent race, while Nieken and Sliwka (2010) also tested pure-strategy versus mixed-strategy equilibria in a patent race setting with a leader and a follower (see Chapter 5.2.2). These studies showed consistency between theory and experiment and they have important implications for competitive behavior; their consistency may reside in the fact that they based their experiments on proven concepts such as Schumpeterian competition or results from game theory, such as mixed-strategy and pure-strategy equilibria.

I believe that the experimental study from this dissertation brings an important contribution to the last three studies I have mentioned. It is also based on a strong game-theoretic foundation but delivered to experimental participants in a very intuitive way, which led to consistency between theory and experiment. Similarly to Amaldoss and Jain (2002) as well as Nieken and Sliwka (2010), the present study is of high relevance for competitive behavior on asymmetric markets with a market leader and a challenger. My contribution goes beyond Nieken and Sliwka, as I introduce a time framework with interest rates as well as a variation of the probability of success for the risky strategy. My experiment also sheds additional light on patent races compared to Amaldoss and Jain: these authors modeled a mixed-strategy equilibrium, while I compared a mixed-strategy and a pure-strategy equilibrium setting. Moreover, I found similar results to other studies of patent races with asymmetric players: the asymmetric position often triggers more aggressive behavior from the side of the disadvantaged market player, while the leader tends to play safe and just protect its current position. In important cases, this can shift the balance of power from the market leader to the challenger, the way Apple destroyed IBM's hegemony by using an aggressive marketing campaign for the introduction of the Macintosh computer in 1984.

Finally, I would like to state that, in my view, the most surprising aspect of this study is subjects' behavior during the experiment. How could they manage to intuitively optimize their decisions in a few minutes, when the mathematical computation of optimal strategies would have taken hours or even days of work? Therefore, my recommendation for future research is to perform more experiments that can analyze human intuition in (economic) decision situations – to the best of my knowledge, important examples of such studies are the publications by Lévesque and Schade (2005), Burmeister-Lamp et al. (2012) or Gigerenzer and Selten (2002). So it was not the Leader, nor the Follower who won or lost the race; on the contrary, Leaders and Followers intuitively harmonized their actions onto optimal play. The real winner of this patent race experiment is the human being.

CHAPTER 7

Conclusions

"Economic theory also pleads for a mechanism design approach: an optimal patent system could be based on a menu of different degrees of patent protection where stronger protection would involve higher fees, allowing self-selection by inventors."

Encaoua et al. (2006)

"Policymakers could increase the inventive step (the standard of nonobviousness in US law) required to obtain a patent so the most trivial advances over the prior art do not qualify for patent protection."

Hunt (1999)

"Empirical studies of innovation have found that end users frequently develop important product and process innovations. Defying conventional wisdom on the negative effects of uncompensated spillovers, innovative users also often openly reveal their innovations to competing users and to manufacturers."

Harhoff et al. (2003)

7. Conclusions

This dissertation analyzed a plethora of aspects related to patents and patent races. The literature review on patents and their effects on economics, business and society pointed out to a hot debate for the case of patent protection. I wrap up the main conclusions about patents in Chapter 7.1. What is more, a review of the patent race literature showed that most of this research branch contains many contradictions and is rather inconclusive. My experimental study on patent races brings an important contribution to the literature by showing that people are able to coordinate their strategies during a patent race, due to two factors: through a wide definition of the equilibrium intervals in the experimental design and because people intuitively play in the direction of equilibrium predictions. This finding is important because it implies that real-world competitors in R&D and innovative activities might also optimize their decisions and avoid socially wasteful behavior. Some empirical evidence seems to confirm this finding. I summarize my analysis of patent races in Chapter 7.2. The last subchapter proposes some solutions to the patent puzzle and discusses potential future developments.

7.1. Conclusions For Patents

In the introductory chapter, the reader was confronted with some of the absurdities of the patent system. The smartphone patent wars brought to light some of the darkest features of patents: false news, continual patent fights, patent thickets, patent trolls, wasteful legal suits, trivial patents impeding innovation, bans of sales for innovative products which restricted consumer choice, as well as unpredictable consequences on firm reputation. The traditional argument for patent protection, still respected today, is that patents are necessary because otherwise innovators who poured great effort into their ideas might be copied by imitators at zero cost. Further arguments in favor of patent protection include patents' important role in fostering information disclosure (as opposed to secrecy) and thus in the diffusion of knowledge. Patents are also essential for the delegation of R&D decisions into the hands of private investors who have the best information on the costs and benefits of R&D (as opposed to a world where the government makes R&D investments and these are funded by taxpayers). While the traditional argument in itself is so strong that it cannot be rejected, I find some situations in which the conventional view is reversed, and I underline many unintended distortions induced by the patent system. In Chapter 1.1.2

I identified some of these drawbacks, most notably the deadweight loss brought by monopoly pricing, the decoupling of patenting and R&D activities, the tendency to generate substitute rather than complementary goods and the significant amount of financial resources that are diverted from innovative activities into patenting or, even worse, into patent litigation trials. Chapter 1.1.3 presented empirical evidence on patents' effectiveness as protection mechanism for firms' returns from innovative activities. In two studies, firms consistently evaluated patents as having a limited role in the protection of innovation, while the primary role is played by such mechanisms as first mover advantage, complexity and secrecy. These findings are corroborated by Encaoua et al. (2006), who point out that "competitive rents, in the absence of patent protection, might be sufficient to compensate innovators in certain circumstances". These authors show that when secrecy works as effective protection mechanism, when first-mover advantages are important or when the cost of imitation is high, there seems to be little need for patents. Another setting under which patents might not be needed is that of sequential and complementary innovation, dealt with in Chapter 1.1.4. Here, Bessen and Maskin (2009) assert that if innovations build upon the preceding ones and are the result of complementary research approaches, imitation may promote innovation, while strong patents may inhibit it. The authors relate to empirical evidence of industries – software, computers and semiconductors – which have been particularly innovative under weak patent protection. In fact, a strengthening of patents in the software industry led to less R&D on the side of patent holders, which is aligned with Bessen and Maskin's hypothesis. Chapter 1.2 was mainly concerned with the increase in patenting activities, the simultaneous decline in patent quality and the emergence of "patent thickets". These phenomena were the result of a shift in firm strategy, from a "single patent" to a "patent portfolio" view, and of a shift to a "more is better" patenting mentality. A solution to these rather negative developments might be to raise patentability standards, and I discuss it in Chapter 7.3. Chapter 1.3 mentioned some of the critical societal and ethical issues raised by patents, especially the patenting of genes or life-forms, patents on lifesaving medicine, corporate attempts to block the distribution of generic medicine, trivial patents blocking innovation, as well as legal asymmetries between small and large firms, with smaller firms unable to fight in case of patent infrigement.

Chapter 2 described the instruments that might be used by policymakers to regulate patent protection – patent breadth, height and length – and then showed the shortcomings of trying to design an optimal regulative structure. The notion of patent breadth was found to embody many heterogeneous concepts. I found Klemperer's (1990) definition to be the most useful: "breadth measures how different competitors' products must be in order not to infringe the patent". Patent height, defined by Scotchmer and Green (1990) as the level of novelty requirements for patenting, might in fact be part of the solution to the explosion of patent applications and their diminishing quality. As I mentioned before, I discuss this in the last subchapter. Patent length, the statutory life of a patent, is the regulatory instrument preferred by patent offices for practical reasons: easiness to define and modify it. Nordhaus (1972) developed a model recommending the shortening or elimination of patents for trivial innovations, but slightly longer patents in the case of radical innovations. But even Nordhaus's conclusion about stronger patents might be reversed under a setting of sequential innovations. In Chapter 2.2 I showed why an optimal patent or patent race regulation cannot be achieved. The literature on this topic is very heterogeneous, it consists of different models that follow a variety of goals and are based on miscellaneous assumptions. However, Chapter 2.2.2 revealed an important conclusion: the strengthening of patent protection did not seem to increase R&D spending or innovative output in the U.S. or in Japan. One study by Hunt (2006) even predicts that stronger patents would have negative effects on innovation.

Chapter 3 dealt with the explosion of patents in Europe and around the world. Not only the number of applications increased dramatically in recent decades, but also the average size of a patent, expressed in the number of its claims and pages. Besides, an extreme phenomenon entitled "mega-applications" was identified; such applications have reached up to 100,000 pages or 20,000 claims, but even larger one have been submitted. At the same time, the increase in patenting was not accompanied by a growth in R&D activity, raising further concerns about patent quality. In Chapter 3.2 five determinants of patent voluminosity were analyzed. First, the fee structure seemed to affect firms' patenting strategies. When faced with additional fees on extra claims, firms tried to limit their number of claims, probably compressing more information in fewer claim categories. Therefore, the fee structure may work as a solution to stop the patent flood, and I discuss it in Chapter 7.3. A second reason for the patent boom is represented by internationalization through the Patent Cooperation Treaty. This procedure provides applicants with 18 additional months before incurring the cost of patenting and applicants do not have to pay fees for excess claims, making the PCT a more attractive way to file for a patent. Third, national practices affected patent voluminosity: countries under common law (such as the Anglo-Saxon countries) register larger patent documents than countries under civil law (such as Continental Europe). Besides, the US patent drafting style by itself was shown to be a significant and large determinant of voluminosity. Fourth, the increased technological complexity that society experienced in recent years required more explanations and more detailed descriptions, leading to larger patent texts. Fifth, the emergence of young, highly competitive industries such as biotechnology, organic chemistry or computers was found to contribute to the patent boom. Chapter 3.3 discussed the effects of the unprecedented surge in patenting. Enormous pressure was placed on patent offices, which have huge backlogs of unexamined applications and must allocate increasingly more resources to search and examine the filings, resulting in questionable patent quality standards. The flood of applications also means more search costs for inventors, patent examiners and the general public; it generates uncertainty and compromises the reliability of the entire patent system. Potential solutions, related to the fee structure, are given in the last subchapter.

In Chapter 4, I examined the role patents play in the pharmaceutical industry. The phenomenon observed by Bessen and Maskin (2009) in the software and computer industries, where patent protection negatively impacted innovation, seems to be even more pronounced in the pharmaceutical industry. Here, I found it questionable whether patents fulfill their expected role of stimulating innovation, as patents seem to produce some anomalies that benefit the industry but harm society. The pharma industry was the most profitable of all US industries during the second half of the 20th Century. At the same time, the pharma industry's innovativeness appears to have slightly declined due to a focus on incremental innovation (also labeled as "me too" drugs), instead of developing radical innovations (new molecular entities with the highest therapeutic value). Similar to the results from Chapter 1.1.3, even for pharmaceuticals (where patents' importance is rated as especially high) there is empirical evidence showing that R&D managers rated different protection instruments for their innovations: secrecy or first-mover advantage were more important than patents. An attempt to encourage generic competition through the 1984 Hatch-Waxman Act remained largely unsuccessful and, paradoxically, might have led to an explosion of industry profits. Pharmaceutical companies make significant marketing and PR efforts to build brand loyalty and strong bonds with medical professionals who prescribe the drugs. This way, they are able to control the market and even increase their prices for brand name drugs, while market penetration for generics remains low. Some recommendations for pharmaceuticals can be found in the last subchapter.

7.2. Conclusions For Patent Races

Chapter 5 gave an extensive overview of the patent race literature. This literature proved to be extremely complex and yet inconclusive. It is enough to take a look at only two initial papers, those by Barzel (1968) and by Kamien and Schwartz (1972), to understand the main problem of the literature: each subsequent study makes departures from previous research, in that the research questions as well as the working assumptions of every model are modified. Even though researchers intended to improve upon the existing knowledge, their attempts led to increasing complexity and decreasing compatibility of their results. The theoretical patent race literature is filled with contradictions. For instance, Lee and Wilde (1980) contradict Loury (1979), while Reinganum (1982) contradicts Kamien and Schwartz (1976). Cockburn and Henderson (1994) provide a good explanation for the instability of the literature: “Conclusions from these types of models are also rather fragile: small changes in the timing of moves, the information structure of the game or the treatment of spillovers can easily reverse or weaken any given theoretical result”. Experimental tests of theoretical two-player multistage patent race models were described in Chapter 5.2.1: Silipo (2005) tested Fudenberg et al. (1983), Zizzo (2002) tested Harris and Vickers (1987) and Breitmoser et al. (2010) tested Hörner (2004). Most of the theoretical predictions failed to withstand experimental tests, rendering little value to the standard patent race literature. On the lines of Breitmoser et al. (2010), I also believe that the predictions of theoretical models are highly sensitive to parameter changes, while real subjects in experiments follow behavioral patterns that are much more stable. The experiment in Chapter 6 corroborates this intuition. Chapter 5.2.2 gave examples of prestigious experimental studies based on own theoretical models and not rooted in the theoretical patent race literature. Here, Isaac and Reynolds (1992) elegantly proved the existence of Schumpeterian competition in a carefully designed experimental setting of an R&D race. Amaldoss and Jain (2002) performed an experimental investigation of a mixed-strategy equilibrium in a patent race context and obtained excellent results. I therefore believe that well-designed experimental settings, based on strong theoretical foundations (rather than the shady predictions of the theoretical patent race literature) may prove to be a fruitful research endeavor, now and in the future.

Following this belief, I set out an experimental study of a patent race as described in Chapter 6. Starting from Nalebuff's (1998) intuitive model of a sailing race (which incorporated many important questions from the patent race literature), I developed a mathematical model of an asymmetric patent race. Here, two firms are entrenched in R&D competition with the goal of attaining a patent. Only the first firm which succesfully completes an innovative project is awarded the patent (the prize of the race). One firm, the Leader, has a time advantage of six months over the other firm, the Follower. Each firm has to choose between a SLOW and a FAST strategy. The SLOW strategy takes two years, but it is sure to lead to a completion of the project. The FAST strategy only takes one year, but it only offers a probability of success in the completion of the project. Each strategy has a cost, and time discounting is accounted for through an interest rate. Based on these parameters, I developed a game-theoretic model and performed an equilibrium analysis. In the absence of a closed-form equilibrium solution, I employed a multi-dimensional graphical solution to select the appropriate parameters for the patent race experiment. I chose the parameters based on two requirements: they had to be easy to process by human participants in an experiment and they had to be stable inside the equilibrium paths. Indeed, the experimental results show that the large breadth of the equilibrium intervals from which the parameters were selected played an essential role: subjects behaved in a manner consistent with predictions. Eventually, the experimental design consisted of eight settings generated by the variation of three parameters: i) whether a subject is a Leader or a Follower; ii) a mixed-strategy equilibrium versus a pure-strategy equilibrium; iii) a setting without interest rate and a setting with 10% interest rate per year. Following Schade's (2005) recommendation that "only students' behavior may be influenced by performance-dependent payments given the average level of payments an experimenter can afford", I ran a laboratory controlled incentive-compatible experiment with 48 subjects, mostly students. Each subject had to play seven rounds of the eight settings, totaling 56 rounds per subject. The order of the eight settings was randomly varied from one round to the next. I formulated two benchmark propositions about how subjects should play, based on the assumption of perfect rationality. I also derived two behavioral hypotheses from the framework of bounded rationality. Hypothesis 1, related to extremeness aversion (Tversky and Simonson, 1992), was that individuals would mix between strategies even if the rational choice is to play a pure strategy. Hypothesis 2, based on anchoring-and-adjustment heuristics (Slovic and Lichtenstein, 1971), stated that individuals' first decision situation would influence their subsequent decisions. In Chapter 6.3, I performed an econometric analysis, the results of which are summarized in the following. In the experiment, people behaved surprisingly close to the normative equilibrium predictions, and most of their deviations from optimal play were small and significant. Therefore, the benchmark propositions were rejected, while Hypothesis 1 was confirmed. The results of random-effects regression models showed that the normative predictions did have a positive and highly significant effect on subjects' observed decisions, while factors such as gender or age were insignificant. Followers deviated more, and Leaders less, from normative predictions. A reasonable explanation for this can be found in the equilibrium predictions, which were more unstable for the Follower than for the Leader. There was no evidence for anchoring-and-adjustment heuristics and thus Hypothesis 2 was rejected. Instead, each subject followed an individual tendency to play the FAST or SLOW strategy and subjects discovered their tendencies during play. I believe that the high consistency between theory and experiment puts my study on a par with prestigious publications such as those by Isaac and Reynolds (1992) or Amaldoss and Jain (2002). Implications for patent races are discussed in the next chapter.

7.3. Recommendations and Future Outlook

In this chapter I provide potential solutions to the patent conundrum. I found two sets of solutions, related to patent fees and to the reinforcement of high patentability standards. I also give recommendations for pharmaceuticals as well as patent races and briefly mention future developments.

Encaoua et al. (2006) discuss restructuring the patent fee system as a feasible solution to the patent problem. They write that "economic theory also pleads for a mechanism design approach: an optimal patent system could be based on a menu of different degrees of patent protection where stronger protection would involve higher fees, allowing self-selection by inventors" (idem). The authors suggest that "patent fees should reflect the cost of patents to society, rather than patent offices' examination costs" (idem). They make a series of further recommendations, which I find highly relevant: i) patent application and renewal fees might be useful as a self-selection process "to encourage high valuable inventions to be patented and discourage the least valuable ones"; ii) the funding structure of patent offices should be reformed; iii) governments should regard patent offices "not as profit centres, but as agencies in charge of aspects of innovation policy" (idem).

In the matter of fees, Encaoua et al.'s recommendations are aligned with those of other researchers who suggested that the fee structures enforced by patent offices significantly affect firms' patenting behavior. In Chapter 3.3, Dack and Cohen (2001) proposed a number of useful measures: a more rigorous approach to evaluate the clarity of an application, the introduction of a claims-based fee for PCT applications, and a limitation on the number of claims. Since firms clearly react to additional fees for extra claims (as seen in Chapter 3.2.1), I am inclined to recommend a progressive fee on each additional claim. However, as van Zeebroeck et al. (2008) rightfully show, the delicate issue in setting claim-based fees is to "ensure that a point is not crossed when it becomes cheaper to divide an application than pay excess claim fees". Therefore, any modification of the fee structure must be designed intelligently and take into account a series of arising trade-offs. Encaoua et al. (2006) also note that how an optimal fee structure "can be implemented remains to be carefully investigated".

Hunt (1999) adds to the discussion on fees by saying that "there may be instances where raising patent costs can actually induce more R&D". Hunt transforms the discussion of fees into a discussion of patentability standards by suggesting the introduction of a patent tax, or, even better, by increasing requirements to obtain a patent. In Hunt's (1999) words, "policymakers could increase the inventive step (the standard of nonobviousness in US law) required to obtain a patent so the most trivial advances over the prior art do not qualify for patent protection".

The literature uniformly agrees that society should get rid of "trivial patents". Patents with low social value, or illegitimate patents (not novel or not sufficiently inventive), can negatively impact innovation and competition. Thus, we need to reform the patent system such that patentability standards are "kept high" and are "strictly applied" (Encaoua et al., 2006). Chapter 2 showed one instrument that allows patent policy makers to raise patent quality standards: through patent height regulation. Encaoua et al. (2006) write that preserving a high patent quality is especially important in our newest and most sensitive industries, "notably software and biotechnology, where there is not yet an established examination tradition in patent offices, to allow monitoring of granting standards". Bessen (2003) suggests that many distortions such as trivial patents or patent thickets are the result of low patent standards. Bessen (2003) writes: "When patent standards are high, portfolio building is costly, and firms are more likely to follow a strategy where they do not assert their patents". At the opposite end, low patent standards result in aggressive portfolio building, "socially wasteful behavior" and reduced R&D incentives (idem). Therefore, it should be one of society's main concerns to reestablish high standards for patents (in specialized terminology, to increase patent height).

I also wrap up my recommendations for pharmaceuticals. I suggest that patent protection and its effects on innovativeness should be more carefully analyzed. The record profits achieved in the pharma industry should be more highly taxed. On the lines of Fershtman and Markovich (2010), small entrepreneurial firms should be encouraged, as they were found to be relatively more innovative than big established pharma companies. In the US, payers of insurances should incentivize medical professionals to prescribe more generic medicine, thus loosening the grip that expensive brand name drugs have on the market. Last but not least, the ethical aspects of pharmaceuticals should be emphasized – these are not regular products, but instead they save lives. Attempts such as those undertaken by Gilead Sciences (see Chapter 1.3) to prevent the distribution of generic lifesaving drugs should be stopped. One instrument that could be employed in such situations is compulsory licensing, where the patent holder is forced to license its patent under fair and reasonable conditions.

For patent races, I suggest the implementation of further experiments based on stable theoretical foundations and intuitive designs, with the goal of teasing out people's behavioral patterns or their intuition in competitive settings. My experiment showed that subjects were able to coordinate their strategies onto optimal play. Very good results were also found in three experiments (Isaac and Reynolds, 1992, Amaldoss and Jain, 2002, Nieken and Sliwka, 2010). I believe that future research should take into account the elegant designs of these authors, as well as my experiment, and should build new experiments based on established and solid theoretical foundations. In the literature, I only found one empirical example of real racing behavior, namely in the pharmaceutical industry (on which Chapter 4.4 provides a detailed investigation). Here, firms avoided cut-throat competition or wasteful duplication of costs and managed to share the market equitably. Therefore my experimental results on harmonized strategies that maximize profits for everyone might also reflect an industry reality. What is more, given that empirical evidence cannot be easily to the theoretical constructs of the standard patent race literature, tremendous importance can be assigned to experiments, as they remain the only way to verify the theoretical findings of patent race models.

Society seems to be advancing towards a new paradigm for innovation, and criticism of intellectual property has lately intensified. The dramatic increase of communication and connectivity from recent years has paved the way for a new phenomenon: open innovation. In traditional, closed-innovation settings, firms keep strict control of R&D, development and production, try to prevent knowledge spillovers to competitors, deliver products down the pipeline to passive consumers and spend a lot on patenting. In an open-innovation framework collaboration is increased, knowledge is shared, and consumers are actively involved in the design of new, better products. In such a setting, it becomes harder and harder to protect intellectual property.

However, open innovation cannot function by itself without the incentives to innovate provided by intellectual property protection. The main argument in favor of patents has always been, and continues to be respected: without protection from imitators, true innovators have no reason to bring innovations to the market, because they cannot hope for any appropriability of benefits from their efforts. Thus, phenomena such as open innovation should not and cannot go too far as to destroy incentives to innovate altogether. I trust that an equilibrium shall be found between a setting with too much and too little intellectual property protection.

The smartphone patent wars were very different from previous patent fights (as emphasized by The Economist, 21.10.2010), because a new type of player arrived on the battlefield, embodied by firms with open-source platforms. Google, for instance, does not charge for its operating system for smartphones, Android, and lets others modify the software. This threatens vendors of closed-innovation proprietary platforms, such as Apple and Microsoft, but benefits society by encouraging innovation. Even though some of the established firms will fight to protect the *status quo*, in the coming decades I expect a shift towards open innovation, resulting in more collaboration, a changing role or new design for patents and ultimately a better society.

Bibliography

Amaldoss, W. and Jain, S. (2002): David vs. Goliath: An analysis of asymmetric mixed-strategy games and experimental evidence. Management Science, 48, 972-991.

Angell, M. (2004): The truth about the drug companies: How they deceive us and what to do about it. New York, NY: Random House.

Aoki, R. (1991): Competition for product innovation: an endless race. American Economic Review, 81, 252–256.

Archontopoulos, E., Guellec, D., Stevnsborg, N., van Pottelsberghe de la Potterie, B. and van Zeebroeck, N. (2007): When small is beautiful: measuring the evolution and consequences of the voluminosity of patent applications at the EPO. Information Economics and Policy, 19, 103–132.

Ariely, D. (2008): Predictably irrational. Harper Collins, London.

Arrow, K. (1962): The economic implications of learning by doing. The Review of Economic Studies, 29, 155-173.

Baker, R. (1976): New and improved: Inventors and inventions that have changed the modern world. London: British Museum Publications.

Barzel, Y. (1968): Optimal timing of innovations. The Review of Economics and Statistics, 50, 348-355.

Baye, M.R. and Hoppe, H.C. (2003): The strategic equivalence of rent-seeking, innovation, and patent-race games. Games and Economic Behavior, 44, 217-226.

Beath, J., Katsoulacosy, Y. and Ulph, D. (1987): Sequential product innovation and industry evolution. Economic Journal (Conference Papers), 97, 32-43.

Bessen, J. (2003): Patent thickets: strategic patenting of complex technologies. SSRN Working Paper. Available at SSRN: http://ssrn.com/abstract=327760.

Bessen, J. and Hunt, R.M. (2004): An empirical look at software patents. Federal Reserve Bank of Philadelphia, Working Paper No. 03-17.

Bessen, J. and Maskin, E. (2009): Sequential innovation, patents, and imitation. RAND Journal of Economics, 40, 611-635.

Blind, K., Edler, J., Frietsch, R. and Schmoch, U. (2003): Erfindungen kontra Patente, Studien zum deutschen Innovationssystem. http://www.technologische-leistungsfaehigkeit.de (Accessed: 12.08.2013).

Boldrin, M. and Levine, D.K. (2002): The case against intellectual property. American Economic Review, Papers and Proceedings, 92, 209–212.

Breitmoser, Y., Tan, J.H.W. and Zizzo, D.J. (2010): Understanding perpetual R&D races. Economic Theory, 44, 445–467.

Bronars, S. (1986): Strategic behavior in tournaments. Austin, Texas A&M University.

Burmeister-Lamp, K., Lévesque, M. and Schade, C. (2012): Are entrepreneurs influenced by risk attitude, regulatory focus or both? An experiment on entrepreneurs' time allocation. Journal of Business Venturing, 27, 456-476.

Cable News Network (2001): Baycol linked to 52 deaths. http://europe.cnn.com/2001/BUSINESS/08/13/bayer/ (Accessed 06.07.2012)

Camerer, C. and Ho, T. (1999): Experience-weighted attraction learning in games. Econometrica, 87, 827-874.

Camerer, C.F. (2011): The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. Working Paper Series.

Cantner, U., Güth, W., Nicklisch, A. and Weiland, T. (2004): Competition in innovation and imitation - A theoretical and experimental study. Jenaer

Schriften zur Wirtschaftswissenschaft 01/2004, Friedrich-Schiller-Universität Jena, Wirtschaftswissenschaftliche Fakultät.

Cantner, U., Nicklisch, A. and Weiland, T. (2005): Innovation races: An experimental study on strategic research activities. Jenaer Schriften zur Wirtschaftswissenschaft 17/2005, Friedrich-Schiller-Universität Jena, Wirtschaftswissenschaftliche Fakultät.

Caves, R.E., Whinston, M.D. and Hurwitz, M.A. (1991): Patent expiration, entry, and competition in the U.S. pharmaceutical industry. Brookings Papers on Economic Activity: Microeconomics, 1-66.

Chong, P. H., Seeger, J.D. and Franklin, C. (2001): Clinically relevant differences between the statins: Implications for therapeutic selection. The American Journal of Medicine, 111, 390–400.

Chu, A.K. and Peng, S. (2011): International intellectual property rights: Effects on growth, welfare and income inequality. Journal of Macroeconomics, 33, 276–287.

Cockburn, I. and Henderson, R. (1994): Racing to invest? The dynamics of competition in ethical drug discovery. Journal of Economics and Management Strategy, 3, 1-19.

Cohen, W.M. and Levin, R.C. (1989): Empirical studies of innovation and market structure. In: Schmalensee, R. and Willig, R.D. (eds.) The Handbook of Industrial Organization, 1059–1107. New York: North Holland.

Cohen, W.M., Nelson, R.R. and Walsh, J.P. (2000): Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not). Working Paper 7552. Cambridge, MA: National Bureau of Economic Research.

Crawford, V.P. (1997): Theory and experiment in the analysis of strategic interaction. Advances in Economics and Econometrics: Theory and Applications, 206-242.

Cremers, K. and Licht, G. (2006): Anstieg der Patentanmeldungen – Fakten und Erklärungsversuche. Die Volkswirtschaft, 7/8-2006, 26-29.

Dack, S. and Cohen, B. (2001): Complex applications – A return to first principles. IIC International Review of Industrial Property and Copyright Law, 32, 485–504.

Dasgupta, P. and Stiglitz, J. (1980): Uncertainty, industrial structure and the speed of R and D. Bell Journal of Economics, 11, 1-28.

Delbono, F. (1989): Market leadership with a sequence of history dependent patent races. The Journal of Industrial Economics, 38, 95-101.

Deneckere, R. and Peck, J. (1995): Competition over price and service rate when demand is stochastic: A strategic analysis. RAND Journal of Economics, 26, 148-162.

Denicolò, V. (1996): Patent races and optimal patent breadth and length. The Journal of Industrial Economics, 44, 249-265.

DiMasi, J.A., Hansen, R.W., Grabowski, H.G. and Lasagna, L. (1991): The cost of innovation in the pharmaceutical industry. Journal of Health Economics, 10, 107-142.

DiMasi, J.A. and Pacquette, C. (2004): The economics of follow on drug research and development: trends in the entry rates and timing of development. Pharmacoeconomics, 22, 1–14.

DiMasi, J.A. and Faden, L. B. (2011): Competitiveness in follow-on drug R&D: a race or imitation? Nature, 10, 23-27.

Duncan, J.R. (1965): The European patent convention as a guide to modernizing our patent examination system. Journal of the Patent Office Society, 47, 220–297.

Encaoua, D., Guellec, D., Martinez, C. (2006): Patent systems for encouraging innovation: Lessons from economic analysis. Research Policy, 35, 1423–1440.

Eswaran, M. and Gallini, N. (1996): Patent policy and the direction of technical change. The Rand Journal of Economics, 27, 722–746.

ETC Group (2008): Who owns nature? Corporate power and the final frontier in the commodification of life. ETC Group Communiqué No. 100.

Farrell, J. and Saloner, G. (1988): Coordination through committees and markets. RAND Journal of Economics, 19, 235-252.

Fershtman, C. and Markovich, S. (2010): Patents, imitation and licensing in an asymmetric dynamic R&D race. International Journal of Industrial Organization, 28, 113–126.

Field, A. (2009): Discovering statistics using SPSS. Third Edition, Sage Publications.

Fischbacher, U. (2007): Z-tree: Zurich tooldbox for ready-made economic experiments. Experimental Economics 10, 171-178.

Fraunhofer Technologie-Entwicklungsgruppe (1999): Studie zum Technologie- und Patentmanagement.

Freeman, R. (1982): A model of international competition in research and development. Manuscript, Federal Reserve Board.

Fudenberg, D., Gilbert, R., Stiglitz, J. and Tirole, J. (1983): Preemption, leapfrogging and competition in patent races. European Economic Review, 1, 3-31.

Fudenberg, D. and Tirole, J. (1985): Preemption and rent equalization in the adoption of new technology. Review of Economic Studies, 52, 383-401.

Futia, C. A. (1980): Schumpeterian Competition. Quarterly Journal of Economics, 95, 675-695.

Gigerenzer, G. and Selten, R. (2002): Bounded rationality: The adaptive toolbox. MIT Press.

Gilbert, R. and Newberry, D. (1982): Preemptive patenting and the persistence of monopoly. American Economic Review, 72, 514–526.

Gilbert, R. and Shapiro, C. (1990): Optimal patent length and breadth. RAND Journal of Economics, 21, 106-112.

Grabowski, H.G. and Wang, Y.R. (2006): The quantity and quality of worldwide new drug introductions, 1982-2003. Health Affairs, 25, 452-460.

Griliches, Z. (1990): Patent statistics as economic indicators: A survey. Journal of Economic Literature, 28, 1661–1707.

Grossman, G.M. and Shapiro, C. (1987): Dynamic R&D programs. Economic Journal, 97, 372–387.

Gruber, H. (1994): Learning and strategic product innovation: Theory and evidence for the semiconductor industry. Amsterdam, North Holland.

Hagedoorn, J., Link, A. N. and Vonortas, N. S. (2000): Research partnerships. Research Policy, 29, 567–586.

Hall, B. and Ziedonis, R. (2001): The patent paradox revisited: an empirical study of patenting in the U.S. semiconductor industry 1979–1995. RAND Journal of Economics, 32, 101–128.

Harhoff, D. (1997): Innovationsanreize in einem strukturellen Oligopolmodell, Discussion Paper FS IV 97 - 9, Wissenschaftszentrum Berlin.

Harhoff, D., Henkel, J. and von Hippel, E. (2003): Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. Research Policy, 32, 1753–1769.

Harris, C. and Vickers, J. (1985): Perfect equilibrium in a model of a race. Review of Economic Studies, 52, 193–209.

Harris, C. and Vickers, J. (1987): Racing with uncertainty. The Review of Economic Studies, 54, 1-21.

Harsanyi, J. (1973): Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. International Journal of Game Theory, 2, 1-23.

Hartwick, J.M. (1991): Patent races optimal with respect to entry. International Journal of Industrial Organization, 9, 197-207.

Heller, M. A. and Eisenberg, R.S. (1998): Can patents deter innovation? The anticommons in biomedical research. Science, 280, 698-701.

Hollis, A. (2005): Comment on “The economics of follow-on drug research and development: trends in entry rates and the timing of development”. Pharmacoeconomics 3, 1187–1192.

Holt, C.A. and Laury, S.K. (2002): Risk aversion and incentive effects. The American Economic Review, 92, 1644-1655.

Hopenhayn, H., Llobet, G. and Mitchell, M. (2006): Rewarding sequential innovators: Prizes, patents, and buyouts. Journal of Political Economy, 114, 1041-1068.

Hörner, J. (2004): A perpetual race to stay ahead. Review of Economic Studies, 71, 1065–1088.

Horstmann, I., MacDonald, G.M. and Slivinski, A. (1985): Patents as information transfer mechanisms: To patent or (maybe) not to patent. Journal of Political Economy, 93, 837–858.

Hunt, R.M. (1999): Nonobviousness and the incentive to innovate: An economic analysis of intellectual property reform. Federal Reserve Bank of Philadelphia, Working Paper No. 99-3.

Hunt, R.M. (2006): When do more patents reduce R&D? The American Economic Review, 96, 87-91.

Hunter, J.E. (2001): The desperate need for replications. Journal of Consumer Research, 28, 149-158.

Huttin, C. (2002): Experiences with budgets and experiences with reference prices. International Journal of Risk & Safety in Medicine, 15, 81-91.

Isaac, R.M. and Reynolds, S.S. (1992): Schumpeterian competition in experimental markets. Journal of Economic Behavior and Organization, 17, 59-100.

Jaffe, A.B. (2000): The U.S. patent system in transition: policy innovation and the innovation process. Research Policy, 29, 531–557.

Judd, K.L., Schmedders, K. and Yeltekin, S. (2012): Optimal rules for patent races. International Economic Review, 53, 23-52.

Kähkönen, A. (2005): Patent race experiment revisited: Are differences between experiments just random variation? Mimeo, University of Joensuu, Department of Business and Economics, Finland.

Kamien, M. I. and Schwartz, N.L. (1972): Timing of innovations under rivalry. Econometrica, 40, 43-60.

Kamien, M. I., Schwartz, N. L. (1974): Patent life and R&D rivalry. American Economic Review, 64, 183-87.

Kamien, M.I. and Schwartz, N.L. (1976): On the degree of rivalry for maximum innovative activity. Quarterly Journal of Economics, 90, 245-260.

Kesselheim, A.S. and Mello, M.M. (2007): Confidentiality laws and secrecy in medical research: Improving public access to data on drug safety. Health Affairs, 26, 483-491.

Klein, N. (2009): The shock doctrine: The rise of disaster capitalism. Knopf Canada.

Klemperer, P. (1990): How broad should the scope of patent protection be? RAND Journal of Economics, 21, 113-130.

Knoeber, C.R. and Thurman, W.N. (1994): Testing the theory of tournaments: An empirical analysis of Broiler production. Journal of Labor Economics, 12, 155-179.

Kortum, S. and Lerner, J. (1999): What is behind the recent surge in patenting? Research Policy, 28, 1-22.

Kultti, K., Takalo, T. and Toikka, J. (2007): Secrecy versus patenting. RAND Journal of Economics, 38, 22-42.

Lan, L. and Kanafani, A. (1993): Economics of park-and-shop discounts: A case of bundled pricing strategy. Journal of Transport Economics and Policy, 27, 291-303.

Lanjouw, J. and Schankerman, M. (1997): Stylized facts of patent litigation: Value, scope and ownership. NBER Working Paper 6297.

Lanjouw, J. and Schankerman, M. (1999): The quality of ideas: measuring innovation with multiple indicators. NBER Working Paper 7345.

Lanjouw, J.O. and Schankerman, M. (2004): Patent quality and research productivity: Measuring innovation with multiple indicators. The Economic Journal, 114, 441–465.

Lee, T. and Wilde, L.L. (1980): Market structure and innovation: A reformulation. Quarterly Journal of Economics, 94, 429-436.

Lerner, J. (1997): An empirical exploration of a technological race. RAND Journal of Economics, 28, 228–247.

Lerner, J. (2009): The empirical impact of intellectual property rights on innovation: puzzles and clues. American Economic Review: Papers and Proceedings, 99, 343-348.

Lévesque, M. and Schade, C. (2005): Intuitive optimizing: experimental findings on time allocation decisions with newly formed ventures. Journal of Business Venturing, 20, 313–342.

Levin, R.C., Klevorick, A.K., Nelson, R.R. and Winter, S.G. (1987): Appropriating the returns from industrial research and development. Brookings Papers on Economic Activity, 3, 783–831.

Licht, G. and Zoz, K. (1998): Patents and R&D, an econometric investigation using applications for German, European and US patents by German companies. Annales d'Economie et de Statistique, 49/50, 329–360.

Light, D.W. and Warburton, R. (2011): Demythologizing the high costs of pharmaceutical research. BioSocieties, 6, 34–50.

Loury, G. C. (1979): Market structure and innovation. Quarterly Journal of Economics, 93, 395-410.

Lu, Z.J. and Comanor, W.S. (1998): Strategic pricing of new pharmaceuticals. Review of Economics and Statistics, 80, 108-118.

Magazzini, L., Pammolli, F. and Riccaboni, M. (2004): Dynamic competition in pharmaceuticals: Patent expiry, generic penetration, and industry structure. The European Journal of Health Economics, 5, 175-182.

Mansfield, E. (1963): Size of firm, market structure, and innovation. Journal of Political Economy, 71, 556-76.

Mansfield, E. (1984): R&D and innovation: Some empirical findings. In Griliches, Z. (ed.): R&D, patents and productivity, Chicago: University of Chicago Press for the National Bureau of Economic Research, 127-154.

Maskin, E. and Tirole, J. (2001): Markov perfect equilibrium. Journal of Economic Theory, 100, 191-219.

Matutes, C., Regibeau, P. and Rockett, K.E. (1996): Optimal patent protection and the diffusion of innovation. RAND Journal of Economics, 27, 60-83.

Meron, A. and Caves, R.E. (1991): Rivalry among firms in research and development outlays. Mimeo, Harvard University.

Mortensen, D. (1982): Property rights and efliciency in mating, racing, and related games. American Economic Review, 72, 968-79.

Nalebuff, B. (1988): Puzzles: Penny stocks, discount brokers, better bidding, and more. The Journal of Economic Perspectives, 2, 179-185.

Nash, J. (1951): Non-cooperative games. The Annals of Mathematics, Second Series, 54, 286-295.

Nelson, R.R., Peck, M.J. and Kalachek, E.D. (1967): Technology economic growth and public policy. Washington, D.C.: Brookings Institution, 53-56.

Nieken, P. and Sliwka, D. (2010): Risk-taking tournaments – Theory and experimental evidence. Journal of Economic Psychology, 31, 254–268.

NIHCM, National Institute for Health Care Management Foundation (2000): Prescription drugs and intellectual property protection: Finding the right balance between access and innovation. Washington D.C.: NIHCM Foundation.

NIHCM, National Institute for Health Care Management Foundation (2002a): Changing patterns of pharmaceutical innovation. Washington D.C.: NIHCM Foundation.

NIHCM, National Institute for Health Care Management Foundation (2002b): A primer: generic drugs, patents and the pharmaceutical marketplace. Washington D.C.: NIHCM Foundation.

Nordhaus,W.D. (1969): Invention growth, and welfare: a theoretical treatment of technological change. MIT Press, Cambridge.

Nordhaus, W.D. (1972): The optimum life of a patent: Reply. American Economic Review, 62, 428-431.

O'Donoghue, T., Scotchmer, S. and Thisse, J. (1998): Patent breadth, patent life and the pace of technological progress. Journal of Economics and Management Strategy, 7, 1–32.

Okamoto, T., Kitakaze, T., Kitamura, Y., Kokaji, N., Kondo, M., Jimbo, K., Tamura, K. and Nakamura, T. (1996): Total patent costs (filing and maintenance costs). Working Paper, presented at the 27th Pacific Intellectual Property Association General Meeting in Hiroshima.

Oshinsky, D.M. (2005): Polio: An American story. Oxford University Press, Oxford.

Park, J. (1987): Dynamic patent races with risky choices. Management Science, 33, 1563-1571.

Patchett, A.A., Harris, E., Tristram, E.W., Wyvratt, M.J., Wu, M.T., Taub, D., Peterson, E.R., Ikeler, T.J., Broeke, J., Payne, L.G., Ondeyka, D.L., Thorsett, E.D., Greenlee, W.J., Lohr, N.S., Hoffsommer, R.D., Joshua, H., Ruyle, W.V., Rothrock, J.W., Aster, S.D., Maycock, A.L., Robinson, F.M. and Hirschmann, R. (1980): A new class of angiotensin-converting enzyme inhibitors. Nature, 288, 280-282.

PhRMA, Pharmaceutical Research and Manufacturers of America (2011): Drug patents are good for our health. http://web.archive.org/web/20110817141824/http://www.phrma.org/drug-patents-are-good-our-health

PhRMA, Pharmaceutical Research and Manufacturers of America (2012): PhRMA 2012 Profile. PhRMA, 950 F STREET, NW WASHINGTON, DC 20004.

Pompidou, A. (2005): Opening statement to the EPO Conference on patent quality in the patent system. The Hague.

Rasmusen, E. (2001): Games and information, Third Edition. An introduction to game theory, Oxford: Blackwell Publishing.

Reinganum, J. F. (1982): A dynamic game of R and D: Patent protection and competitive behavior. Econometrica, 50, 671-688.

Reinganum, J.F. (1984): Practical implications of game theoretic models of R&D. American Economic Review, 74, 61-66.

Reinganum, J. F. (1985): Innovation and industry evolution. Quarterly Journal of Economics, 99, 81-99.

Reinganum, J. F. (1989): The timing of innovation: Research, development and diffusion. In: Schmalensee, R., Willig, R.D. (Eds.), Handbook of Industrial Organization, vol. I, Elsevier, Amsterdam.

Sakakibara, M. and Branstetter, L. (2001): Do stronger patents induce more innovation? Evidence from the 1988 Japanese patent law reforms. The RAND Journal of Economics, 32, 77-100.

Sbriglia, P. and Hey, J.D. (1994): Experiments in multi-stage R&D competition. Empirical Economics, 19, 291-316.

Schade, C. (2005): Dynamics, experimental economics, and entrepreneurship. Journal of Technology Transfer, 30, 409-431.

Schade, C. and Burmeister-Lamp, K. (2009): Experiments on entrepreneurial decision making: A different lens through which to look at entrepreneurship. Foundations and Trends in Entrepreneurship, 5, 81–134.

Schade, C. (2013): Marketing of innovations. Lecture material, School of Business and Economics, Humboldt-Universität zu Berlin.

Schankerman, M. (1998): How valuable is patent protection? Estimates by technology field. RAND Journal of Economics, 29, 77-107.

Schelling, T.C. (1960): The strategy of conflict. Harvard University Press, Cambridge.

Scherer, F.M. (1967a): Research and development allocation under rivalry. Quarterly Journal of Economics, 359-94.

Scherer, F.M. (1967b): Market structure and the employment of scientists and engineers. American Economic Review, 57, 524-531.

Scherer, F.M. (2007): Pharmaceutical innovation. Working Paper No. 07-19. Washington, D.C.: AEI-Brookings Joint Center for Regulatory Studies.

Scherer, F.M. (2009): The political economy of patent policy reform in the United States. Colorado Journal on Telecommunications and High Technology Law, 7, 167-216.

Schmidt, R.H. and Terberger, E. (1997): Grundzüge der Investitions- und Finanzierungstheorie. 4. Auflage, Wiesbaden: Gabler.

Schondelmeyer, S.W. (1999): Patent extensions of pipeline drugs: Impact on U.S. health care expenditures. PRIME Institute, College of Pharmacy, University of Minnesota.

Schumpeter, J. A. (1942): Capitalism, socialism and democracy. New York, NY: Harper and Row.

Schumpeter, J.A. (1950): Capitalism, socialism and democracy, third ed. Harper & Brothers, New York.

Scotchmer, S. and Green, J. (1990): Novelty and disclosure in patent law. RAND Journal of Economics, 21, 131-146.

Scotchmer, S. (2004): Innovation and incentives. Cambridge, MA: The MIT Press.

Shapiro, C. (2001): Navigating the patent thicket: Cross licenses, patent pools, and standard-setting. NBER Working Paper.

Siebert, R. and von Graevenitz, G. (2010): Jostling for advantage or not: Choosing between patent portfolio races and ex ante licensing. Journal of Economic Behavior and Organization, 73, 225–245.

Silipo, D.B. (2005): The evolution of cooperation in patent races: Theory and experimental evidence. Journal of Economics, 85, 1-38.

Simon, H. (1957): A behavioral model of rational choice. In: Models of man, social and rational: Mathematical essays on rational human behavior in a social setting. New York, Wiley.

Simons, J. (2003): The $10 billion pill. Fortune magazine, 147, 58.

Slovic, P. and Lichtenstein, S. (1971): Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.

Smith, A.L. (2003): The history of the patent and trademark office society. Supplement to Journal of the Patent and Trademark Office Society, 85th Anniversary Edition: 1917–2002, Washington DC, USA, 13–22.

Smith, V.L. (1976): Experimental economics: Induced value theory. American Economic Review 66, 274-279.

Spence, M. (1982): Cost Reduction, Competition and Industry Performance. Harvard Institute of Economic Research, Discussion Paper.

Stevnsborg, N. and van Pottelsberghe de la Potterie, B. (2007): Patenting procedures and filing strategies. In Guellec, D. and van Pottelsberghe de la Potterie, B.: The economics of the European patent system, Oxford University Press, Oxford, 155–183.

Stewart, M. (1983): Noncooperative oligopoly and preemptive innovation without winner-take-all. Quarterly Journal of Economics, 98, 681-94.

Taylor, C.T. and Silberston, Z.A. (1973): The economic impact of the patent system - A study of the British experience. Cambridge: Cambridge University Press.

The Economist (13.01.2013): The Supreme Court rules that genes may not be patented. http://www.economist.com.

The Economist (21.10.2010): The great patent battle. http://www.economist.com.

The Guardian (22.10.2012): Apple, Samsung, Google and the smartphone patent wars - everything you need to know. http://www.guardian.co.uk.

The Guardian (23.06.2005): Patent absurdity. http://www.guardian.co.uk.

The Guardian (25.08.2012): Apple awarded over $1bn in Samsung patent infringement trial. http://www.guardian.co.uk.

The Guardian (29.08.2012): No, Samsung did not try to pay Apple its $1bn fine in nickels. http://www.theguardian.com.

The New York Times (06.03.2011), p. A1.

The New York Times (22.09.1983), p. B19.

The Telegraph (14.12.2010): Tech body: patents are 'useless'. http://www.telegraph.co.uk.

The Wall Street Journal (22.04.2011): Samsung sues Apple on patents. http://www.wsj.com.

Todd, P.M. and Gigerenzer, G. (2003): Bounding rationality to the world. Journal of Economic Psychology, 24, 143–165.

Tong, X. and Frame, J. (1994): Measuring national technological performance with patent claims data. Research Policy, 23, 133–141.

Tversky, A. and Kahneman, D. (1974): Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131

Tversky, A. and Simonson, I. (1993): Context-dependent preferences. Management Science, 39, 1179-1189.

Ungern-Sternberg, T. (1980): Current profits and investment behavior. Bell Journal of Economics, 11, 745-48.

Van Dijk, T. (1996): Patent height and competition in product improvements. The Journal of Industrial Economics, 44, 151-167.

Van Pottelsberghe de la Potterie, B. and François, D. (2006): The cost factor in patent systems, CEPR discussion papers, 5944.

Van Zeebroeck, N., Stevnsborg, N., van Pottelsberghe de la Potterie, B., Guellec, D. and Archontopoulos, E. (2008): Patent inflation in Europe. World Patent Information, 30, 43–52.

Van Zeebroeck, N., van Pottelsberghe de la Potterie, B. and Guellec, D. (2009): Claiming more: The increased voluminosity of patent applications and its determinants. Research Policy, 38, 1006–1020.

Varian, H. (1980): A model of sales. American Economic Review, 70, 651-659.

Vickers, J. (1986): The evolution of market structure when there is a sequence of innovations. Journal of Industrial Economics, 35, 1-12.

Vonortas, N. S. (2000): Multimarket contact and inter-firm cooperation in R&D. Journal of Evolutionary Economics 10, 243–271.

Williamson, O.E. (1965): Innovation and market structure. Journal of Political Economy, 73, 67-73.

Wissensallmende (2009): Patent-Wettrüsten: Masse statt Klasse. http://www.wissensallmende.de/report-2009/lesen/patent-wettruestenmasse-statt-klasse (Accessed: 05.08.2013).

World Health Organisation (2012): Medicines: spurious/falsely-labelled/falsified/counterfeit (SFFC) medicines. http://www.who.int/mediacentre/factsheets/fs275/en/ (Accessed 06.07.2012).

Zizzo, D.J. (2002): Racing with uncertainty: A patent race experiment. International Journal of Industrial Organization 20, 877-902.

Acknowledgements

I am grateful to my doctoral supervisor, Prof. Dr. Christian D. Schade, for his continuous intensive efforts of advising and coordinating my dissertation, for his indispensible intellectual input, as well as for providing me access to the resources necessary to perform the experimental implementation and the econometric analysis. This dissertation would not have been possible without his vision and guidance.

I am thankful to the Institute for Entrepreneurial Studies and Innovation Management of the Humboldt-Universität zu Berlin for the excellent working conditions and the infrastructure supporting this dissertation. I wish to thank the Head of the Institute, Prof. Dr. Christian D. Schade, and my colleagues Anna Abratis, Dr. Sabrina Artinger, Jessica Baltuttis, Dr. Katrin Burmeister-Lamp, Dr. Wei-Shiun Chang, Tanja Jänicke, Dr. Anna Kostanovskaya, Christine Lauritzen, Ute Ottenbreit, Dr. Hidajet Ramaj, Dr. Serena Sandri and Dr. Avichai Snir. I would like to express additional thanks to Dr. Anna Kostanovskaya for her help with the econometric analysis and to Dr. Avichai Snir for his suggestion of the Nalebuff (1988) model which underlies my experimental investigation.

Last but not least, I am thanking my family and friends for their incessant personal and emotional support.

Appendix

Appendix 6A: Experimental Instructions

(These instructions have been translated from German into English.)

Welcome and thank you for taking part in our experiment on strategic decision behavior.

General Rules

Switch off your cell phone and all other electronic devices.
Please stay absolutely quiet during the entire experiment and do not speak with the other participants.
Concentrate on the given tasks and place yourself in the described situation.
If you have questions, please raise your hand. The experimenter will come to you and explain your question. Please ask your questions quietly, in order not to disturb the other participants.
On some computer screens you will find a calculator. You may use it, but you are not required to.
None of your decisions will be related with your name or personal information.

This experiment consists of two parts. You are now starting the first part of the experiment.

You are the research and development (R&D) manager of a high-tech company and you are presently undertaking a new, costly R&D project.

You have a competitor. Another company is pursuing an R&D project, with the same goal as you. The company which completes the technical solution FIRST (the winner) receives a patent from the patent office. This company is economically successful.

The patent gives the winner the exclusive right to develop a product, based on the relevant results of the R&D project, to sell that product and obtain profits. The company which does not achieve patent protection (the loser) is not allowed to sell any products based on the relevant results of the R&D project, but it must pay its R&D costs. Therefore, this company makes a loss from the project.

Thus, of essential importance is which of the two companies completes the technical solution FIRST. The faster company will become the winner, while the slower company will become the loser.

You and your competitor start with the development of the technical solution in different time points. One of the companies may start immediately with the development of the technical solution. This company is the Leader of the competition. The Leader has an advantage over her competitor, who is named the Follower. The Follower can only start to develop the technical solution in 6 months. Therefore, the Follower has a disadvantage in this time competition.

YOUR DECISION
For the development of the technical solution you may, independent on whether you are the Leader or the Follower, choose one of two different plans: Plan **SLOW** or Plan **FAST**.

The Plan SLOW takes **2 years**, but it is certain to lead you to the completion of the technical solution. The question then remains whether you were the first or the second to complete this technical solution. If you were the first, you receive the patent and you are economically successful. If you were the second, you do not receive the patent and you are not economically successful.

The Plan FAST only takes **1 year**, but it is not certain that you can develop the technical solution.
Sometimes, the Plan FAST offers a chance to complete the technical solution of **1 out of 3** (probability one third).
Sometimes, the Plan FAST offers a chance to complete the technical solution of **1 out of 2** (probability one half).
Thus it is possible that you do not complete the technical solution with Plan FAST. In this case, you have to switch to Plan SLOW, pay again the R&D costs, and take another 2 years to complete the technical solution, which however might be too late to receive the patent.

The competition starts on the *1st of January, Year 1*. In this time point you receive an initial endowment amounting to **600.000** Talers.
The R&D costs for the Plans SLOW or FAST each amount to **250.000** Talers. These costs will be subtracted from your initial endowment.
Receiving the patent brings you a revenue of **1.000.000** Talers.
Case 1: Sometimes, your revenues and costs are not subject to an interest rate. This will be communicated to you at the beginning of the decision situation. In this case, the value of the individual revenues and costs, and their sum remain unchanged over three years; that is, until the *31st of December, Year 3*.
Case 2: Sometimes, your revenues and costs are subject to an interest rate of **10%** per year. This will also be communicated to you at the beginning of the decision situation. In this case, the value of the individual revenues and costs will be compounded with the interest rate over the three years, until the *31st of December, Year 3*.
Thus, the *31st of December, Year 3* represents the liquidation time point for your revenues and costs.
In the following you are shown the effects of the interest rate on your revenues and costs.

Let's suppose you are the **Leader**. You receive your initial endowment of **600.000** Talers on the *1st of January, Year 1*. On the same day you have to make your decision between Plan SLOW and Plan FAST, and you also pay the R&D costs of **250.000** Talers.

Let's assume you chose Plan SLOW. Let's further assume the **Follower** has also chosen Plan SLOW, or she failed with Plan FAST. Then, you receive the patent on the *1st of January, Year 3*, and the corresponding revenue of **1.000.000** Talers.

Case 1: no interest rate (will be communicated to you before your decision)

Your payoff on the *31st of December, Year 3* will be:

600.000 - 250.000 + **1.000.000** = **1.350.000** Talers.

Case 2: interest rate of **10%** per year (will be communicated to you before your decision)

Your payoff on the *31st of December, Year 3* will be:

600.000 x 110% x 110% x 110% - **250.000** x 110% x 110% x 110% + **1.000.000** x 110% = **1.565.850** Talers.

(The initial endowment and the costs are subject to the interest rate for 3 years, because they are incurred 3 years before the liquidation date. The revenue from receiving the patent is subject to the interest rate for only 1 year, because it was obtained 1 year before the liquidation date.)

Let's assume you are the **Follower**. You receive your initial endowment of **600.000** Talers on the *1st of January, Year 1*. You have to make your decision 6 months later on whether you choose Plan SLOW or Plan FAST, and you pay the R&D costs of **250.000** Talers on the same day, the *1st of July, Year 1*.

Let's assume you chose Plan SLOW. Let's further assume the **Leader** has chosen Plan FAST, and failed. Then, you receive the patent on the *1st of July, Year 3*, and the corresponding revenue of **1.000.000** Talers.

Case 1: no interest rate (will be communicated to you before your decision)

Your payoff on the *31st of December, Year 3* will be:

600.000 - 250.000 + **1.000.000** = **1.350.000** Talers.

Case 2: interest rate of **10%** per year (will be communicated to you before your decision)

Your payoff on the *31st of December, Year 3* will be:

600.000 x 110% x 110% x 110% - **250.000** x 105% (rounded) x 110% x 110% + **1.000.000** x 105% (rounded) = **1.530.144** Talers.

(The initial endowment is subject to the interest rate for 3 years, because it is received 3 years before the liquidation date. The costs are subject to the interest rate for 2 years and 6 months, because they have to be paid 2 years and 6 months before liquidation. The revenue from receiving the patent is subject to the interest rate for only 6 months, because it was obtained 6 months before the liquidation date.)

The following tables summarize all parameters of this game.

Leader	begins on the 1st of January, Year 1
Follower	begins 6 months later on the 1st of July, Year 1

Initial endowment	600.000

	Duration	Probability of completing the technical solution	Costs
Plan SLOW	2 years	certain	250.000
Plan FAST	1 year	1/3 or 1/2	250.000

Interest rate, Case 1	none
Interest rate, Case 2	10%

Receiving the patent	1.000.000

You will play this competitive game 56 times against the same person. That is, your competitor remains unchanged over the entire duration of the experiment.
In 28 games you will be the Leader, and in 28 games the Follower. At the beginning of every game you receive the initial endowment of **600.000** Talers. In every game, the following parameters might change (this is communicated to you at the beginning of each game):
-whether you are the Leader or the Follower;
-how high is the chance to complete the technical solution using Plan FAST;
-whether an interest rate is applied or not.
In each game, you have to decide whether you want to follow Plan SLOW or Plan FAST first.

At the end of the experiment, the computer will randomly choose 14 games (7 games in which you were the Leader and 7 games in which you were the Follower).
The sum of your earnings in these 14 randomly chosen games will be paid to you at the end of the experiment. Therefore, try to make the best decision in every game.
Please be patient, the results are not shown to you after each game. At the end of the experiment you will receive a table summarizing all your results.
Your competitor will always be the Leader when you are the Follower, and she will always be the Follower when you are the Leader.
Otherwise, you are both subjected to the same conditions: if you have an interest rate, your competitor will also have it. You have the same initial endowments, the same revenue, the same costs to follow a Plan, and the same chance to complete the technical solution through Plan FAST.
The conversion rate is: **1 € = 982.065 Talers**.

Appendix 6B: Hierarchical Regressions on Setting Choice

Table 6.B.1. Random-effects regression. Dependent variable: Setting Decision. Independent variables: Setting Prediction, Female

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.3003                         Obs per group: min =         8
       between = 0.0103                                        avg =       8.0
       overall = 0.2509                                        max =         8

                                                Wald chi2(2)       =    144.29
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

Setting Decision		Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Setting Prediction	***	.5127595	.0427587	11.99	0.000	.4289539	.5965651
Female		.0376984	.0544312	0.69	0.489	-.0689848	.1443816
Constant	*	.1293093	.0542328	2.38	0.017	.0230149	.2356037
sigma_u		.11861571					
sigma_e		.31742843					
rho		.12252569	(fraction of variance due to u_i)				

Significance levels: *** $p < 0.1\%$; * $p < 5\%$. Data coding: Female = (1 if female, 0 if male).

Table 6.B.2. Random-effects regression. Dependent variable: Setting Decision. Independent variables: Setting Prediction, Female, Age

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.3003                         Obs per group: min =         8
       between = 0.0312                                        avg =       8.0
       overall = 0.2545                                        max =         8

                                                Wald chi2(3)       =    145.25
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

Setting Decision	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Setting Prediction	*** .5127595	.0427587	11.99	0.000	.4289539	.5965651
Female	.0442697	.0548584	0.81	0.420	-.0632507	.1517902
Age	.0067591	.0068704	0.98	0.325	-.0067066	.0202247
Constant	-.0475527	.1877806	-0.25	0.800	-.4155959	.3204905
sigma_u	.11869426					
sigma_e	.31742843					
rho	.12266812	(fraction of variance due to u_i)				

Significance level: *** p < 0.1%. Data coding: Female = (1 if female, 0 if male).

Table 6.B.3. Random-effects regression. Dependent variable: Setting Decision. Independent variables: Setting Prediction, Female, Age, Female * Role, Risk Attitude (only for consistent subject answers)

```
Random-effects GLS regression                   Number of obs      =       288
Group variable: Subject                         Number of groups   =        36

R-sq:  within  = 0.3265                         Obs per group: min =         8
       between = 0.0690                                        avg =       8.0
       overall = 0.2772                                        max =         8

                                                Wald chi2(5)       =    123.55
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

Setting Decision	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Setting Prediction	*** .5769131	.064872	8.89	0.000	.4497663	.7040599
Female	.0290422	.0767427	0.38	0.705	-.1213708	.1794551
Age	.0067236	.0103048	0.65	0.514	-.0134734	.0269206
Female * Role	.0368275	.0578372	0.64	0.524	-.0765314	.1501863
Risk Attitude	-.0211337	.0202703	-1.04	0.297	-.0608628	.0185954
Constant	.0405986	.3025183	0.13	0.893	-.5523264	.6335236
sigma_u	.14214087					
sigma_e	.32215876					
rho	.16294847	(fraction of variance due to u_i)				

Significance level: *** $p < 0.1\%$. Data coding: Female = (1 if female, 0 if male); Role = (1 if Leader, 0 if Follower); Risk Attitude = (1 for very risk loving, 9 for very risk averse).

Appendix 6C: Hierarchical Regressions on Absolute Deviations from Equilibrium

Table 6.C.1. Fixed-effects regression. Dependent variable: Absolute Deviation. Independent variable: Role

```
Fixed-effects (within) regression               Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.0798                         Obs per group: min =         8
       between =      .                                        avg =       8.0
       overall = 0.0675                                        max =         8

                                                F(1,335)           =     29.06
corr(u_i, Xb)  = 0.0000                         Prob > F           =    0.0000

------------------------------------------------------------------------------
AbsoluteDe~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Role *** -.1532616   .0284305    -5.39   0.000    -.2091864   -.0973369
       _cons |*** .3641232   .0201034    18.11   0.000     .3245784     .403668
-------------+----------------------------------------------------------------
     sigma_u |  .11733051
     sigma_e |  .27856067
         rho |  .15067944   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(47, 335) =     1.42             Prob > F = 0.0434
```

Significance level: *** $p < 0.1\%$. Data coding: Role = (1 if Leader, 0 if Follower).

Table 6.C.2. Hausman test comparing the fixed-effects and the random-effects model

	Coefficients (b) fixed	(B) random	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
Role	-.1532616	-.1532616	6.11e-16	.

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

Table 6.C.3. Random-effects regression. Dependent variable: Absolute Deviation. Independent variables: Role, Success Probability of FAST

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.0000                         Obs per group: min =         8
       between = 0.0000                                        avg =       8.0
       overall = 0.0713                                        max =         8

                                                Wald chi2(2)       =     30.78
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

AbsoluteDeviation	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Role	*** -.1532616	.0284024	-5.40	0.000	-.2089293	-.0975939
SuccessProbabilityFAST	-.219734	.1704144	-1.29	0.197	-.5537401	.1142722
_cons	*** .455679	.0743662	6.13	0.000	.309924	.6014341
sigma_u	.06392253					
sigma_e	.2782856					
rho	.05011826	(fraction of variance due to u_i)				

Significance level: *** p < 0.1%.

Table 6.C.4. Random-effects regression. Dependent variable: Absolute Deviation. Independent variables: Role, Success Probability of FAST, Accumulation

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.0000                         Obs per group: min =         8
       between = 0.0000                                        avg =       8.0
       overall = 0.0726                                        max =         8

                                                Wald chi2(3)       =     31.28
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

AbsoluteDeviation	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Role ***	-.1532616	.0284219	-5.39	0.000	-.2089675	-.0975557
SuccessProbabilityFAST	-.219734	.1705315	-1.29	0.198	-.5539695	.1145016
Accumulation	.2091875	.2842191	0.74	0.462	-.3478717	.7662467
_cons	.2360322	.3075679	0.77	0.443	-.3667899	.8388542
sigma_u	.06381841					
sigma_e	.27847672					
rho	.04989813	(fraction of variance due to u_i)				

Significance level: *** p < 0.1%.

Table 6.C.5. Random-effects regression. Dependent variable: Absolute Deviation. Independent variables: Role, Success Probability of FAST, Accumulation, Female

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.0859                         Obs per group: min =         8
       between = 0.0041                                        avg =       8.0
       overall = 0.0732                                        max =         8

                                                Wald chi2(4)       =     31.47
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

AbsoluteDeviation		Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Role	***	-.1532616	.0284219	-5.39	0.000	-.2089675	-.0975557
SuccessProbabilityFAST		-.219734	.1705315	-1.29	0.198	-.5539695	.1145016
Accumulation		.2091875	.2842191	0.74	0.462	-.3478717	.7662467
Female		-.0170772	.0394527	-0.43	0.665	-.0944031	.0602488
_cons		.24884	.3089962	0.81	0.421	-.3567813	.8544614
sigma_u		.06568867					
sigma_e		.27847672					
rho		.05270921	(fraction of variance due to u_i)				

Significance level: *** p < 0.1%. Data coding: Role = (1 if Leader, 0 if Follower); Female = (1 if female, 0 if male).

Table 6.C.6. Random-effects regression. Dependent variable: Absolute Deviation. Independent variables: Role, Success Probability of FAST, Accumulation, Female, Female * Role

```
Random-effects GLS regression                   Number of obs      =       384
Group variable: Subject                         Number of groups   =        48

R-sq:  within  = 0.0865                         Obs per group: min =         8
       between = 0.0041                                        avg =       8.0
       overall = 0.0737                                        max =         8

                                                Wald chi2(5)       =     31.63
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

-------------------------------------------------------------------------------------------
     AbsoluteDeviation |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------+-------------------------------------------------------------------
                  Role | * -.1295554  .0569095    -2.28   0.023     -.241096   -.0180147
SuccessProbabilityFAST |   -.219734   .1707286    -1.29   0.198    -.5543559    .114888
          Accumulation |   .2091875   .2845477     0.74   0.462    -.3485157   .7668907
                Female |   -.001273   .0513428    -0.02   0.980    -.1019031    .099357
           Female_Role |  -.0316083   .0657135    -0.48   0.631    -.1604044   .0971877
                 _cons |   .2369869    .310329     0.76   0.445    -.3712467   .8452205
-----------------------+-------------------------------------------------------------------
               sigma_u |  .06551774
               sigma_e |  .27879867
                   rho |  .05233489   (fraction of variance due to u_i)
-------------------------------------------------------------------------------------------
```

Significance level: * p < 5%. Data coding: Role = (1 if Leader, 0 if Follower); Female = (1 if female, 0 if male).

Table 6.C.7. Random-effects regression. Dependent variable: Absolute Deviation. Independent variables: Role, Success Probability of FAST, Accumulation, Female, Female * Role, Age, Risk Attitude (only for consistent subject answers)

```
Random-effects GLS regression                   Number of obs      =       288
Group variable: Subject                         Number of groups   =        36

R-sq:  within  = 0.0657                         Obs per group: min =         8
       between = 0.0609                                        avg =       8.0
       overall = 0.0649                                        max =         8

                                                Wald chi2(7)       =     19.51
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0067
```

AbsoluteDeviation	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Role	-.1245157	.0639633	-1.95	0.052	-.2498815	.0008501
SuccessProbabilityFAST	-.0071012	.2022698	-0.04	0.972	-.4035427	.3893403
Accumulation	.0740972	.3371163	0.22	0.826	-.5866387	.7348331
Female	.0205752	.0626964	0.33	0.743	-.1023075	.143458
Female_Role	-.0217428	.0752654	-0.29	0.773	-.1692604	.1257748
Age	-.0035035	.0072693	-0.48	0.630	-.0177511	.0107441
Riskattitude	.0192995	.0142993	1.35	0.177	-.0087266	.0473257
_cons	.2413305	.4220668	0.57	0.567	-.5859053	1.068566
sigma_u	.07925819					
sigma_e	.2860527					
rho	.07129727	(fraction of variance due to u_i)				

Data coding: Role = (1 if Leader, 0 if Follower); Female = (1 if female, 0 if male); Risk Attitude = (1 for very risk loving, 9 for very risk averse).